

---

# **BACHELORARBEIT**

---

Herr  
**Alexander Haase**

**Stories in Data:**

**Das Potential von Daten  
und ihr Einfluss  
auf den Journalismus**

2011

Fakultät Medien

---

## **Bachelorarbeit**

---

# **Stories in Data: Das Potential von Daten und ihr Einfluss auf den Journalismus**

Autor:  
**Herr  
Alexander Haase**

Studiengang:  
**Medienmanagement**

Seminargruppe:  
**MM08w1-B**

Erstprüfer:  
**Prof. Horst Müller**

Zweitprüfer:  
**Dipl.-Ing. Sieglinde Klimant**

Einreichung:  
**Mittweida, 25.7.2011**

Verteidigung/Bewertung:  
**Mittweida, 2011**

# **Bachelor Thesis**

---

**Stories in Data:**

**The Potential of Data  
and their Impact  
on Journalism**

author:

**Herr  
Alexander Haase**

course of studies:

**Media Management**

seminar group:

**MM08w1-B**

first examiner:

**Prof. Horst Müller**

second examiner:

**Dipl.-Ing. Sieglinde Klimant**

submission:

**Mittweida, 25.7.2011**

defence/ evaluation:

**Mittweida, 2011**

## **Bibliographische Angaben**

Haase, Alexander:

Stories in Data: Das Potential von Daten und ihr Einfluss auf den Journalismus

Stories in Data: The Potential of Data and their Impact on Journalism

81 Seiten, Hochschule Mittweida, University of Applied Sciences,

Fakultät Medien, Bachelorarbeit, 2011

## **Abstract**

Die zunehmende Vernetzung der Welt über das Internet führt zu einem stetigen Wachstum der verfügbaren, globalen Datenmasse. Relevante Informationen sind schwieriger und nur noch unter Einsatz von Suchwerkzeugen und Filtersystemen zu finden. An diesem Punkt setzt Datenjournalismus an, der die komplexer gewordene Gesellschaft anhand von Daten präziser, schneller und umfangreicher darstellen und erklären kann. Dabei stehen dem Datenjournalismus technologische, mediale und gesellschaftliche Herausforderungen gegenüber. Neben interessanten Beispielen aus dem Datenjournalismus stellt diese Arbeit die journalistischen Potentiale von Daten und die damit verbundenen Herausforderungen an den Journalismus vor und ermöglicht einen umfangreichen Einblick in das Thema. Die beschriebenen Techniken zur erfolgreichen Arbeit mit Daten von der Aggregation bis zur Präsentation in Datenvisualisierungen wurden in einen Workflow für Datenjournalisten integriert, der dem interdisziplinären Feld des Datenjournalismus gerecht wird.

# Inhaltsverzeichnis

<b>Inhaltsverzeichnis .....</b>	<b>V</b>
<b>Abkürzungsverzeichnis .....</b>	<b>VIII</b>
<b>Abbildungsverzeichnis .....</b>	<b>IX</b>
<b>Tabellenverzeichnis.....</b>	<b>X</b>
<b>1     Einleitung .....</b>	<b>1</b>
1.1 Problemhintergrund .....	1
1.2 Zielsetzung .....	1
1.3 Grenzen der Arbeit.....	2
1.4 Bedingungen der Arbeit .....	2
1.5 Aufbau der Arbeit .....	3
1.6 Begriffserklärungen .....	4
<b>2     Der digitale Wandel .....</b>	<b>5</b>
2.1 Data Overload .....	6
2.2 Internet Killed the Newspaper Star .....	8
2.3 Open Data .....	11
2.3.1   Open Content und die Freiheit der Inhalte.....	12
2.3.2   Open Access und die Freiheit der Wissenschaft .....	13
2.3.3   Open Government und die Freiheit öffentlicher Daten .....	14
<b>3     Daten im Journalismus.....</b>	<b>17</b>
3.1 Das journalistische Potential von Daten .....	18
3.1.1   Datenrecycling .....	18
3.1.2   Beschleunigung von Rechercheprozessen .....	19
3.1.3   Zusammenhänge aufdecken .....	19
3.1.4   Exklusivität .....	20
3.1.5   Leserintegration durch Crowdsourcing .....	20
3.1.6   Verdichtetes Storytelling .....	21
3.2 Daten im Einsatz .....	22
3.2.1   Spesenskandal des Guardian.....	22
3.2.2   Verräterisches Handy von ZEIT ONLINE .....	23

3.2.3	Fluglärm-Karte der Taz .....	24
3.2.4	Status quo im Datenjournalismus .....	25
3.2.5	Die Plagiatsaffäre, ein Selbstläufer.....	26
3.3	Herausforderungen für den Journalismus .....	28
3.3.1	Mut zum Datenjournalismus .....	28
3.3.2	Interdisziplinäre Datenteams aufbauen .....	28
3.3.3	Transparenter werden.....	29
3.3.4	Vertrauen generieren .....	30
3.3.5	Journalisten als Kuratoren von Informationen .....	30
3.3.6	Datenworkflow und -strategien entwickeln .....	31
<b>4</b>	<b>Datenworkflow im Journalismus .....</b>	<b>32</b>
4.1	Start und Vorbereitung .....	33
4.1.1	Hintergrundrecherche .....	34
4.1.2	Zielvorgaben definieren .....	34
4.2	Datenaggregation .....	35
4.2.1	Datenbanken nutzen.....	35
4.2.2	Daten selbst aggregieren.....	36
4.3	Datenaufbereitung .....	37
4.3.1	Datenumwandlung.....	37
4.3.2	Datenbereinigung .....	38
4.4	Visuelle Datenanalyse .....	39
4.4.1	Visualisierungstechniken .....	40
4.4.2	Muster erkennen und interpretieren.....	40
4.4.3	Anwendungsgebiete .....	41
4.5	Datenpräsentation .....	42
4.5.1	Werkzeuge zur Kollaboration und Interaktion integrieren.....	42
4.5.2	Aufbereitete Rohdaten ohne Einschränkung bereitstellen.....	43
4.5.3	Interessante Details einzeln aufarbeiten.....	44
4.5.4	Richtlinien guter Datenvisualisierungen einhalten .....	44
<b>5</b>	<b>Datenvisualisierungen .....</b>	<b>45</b>
5.1	Grundsätze guter Datenvisualisierungen.....	45
5.1.1	Data-Ink .....	46
5.1.2	Lie Factor und Dimensionierung.....	47
5.1.3	Farbe .....	48
5.1.4	Schrift und Beschriftung.....	49

---

5.2 What They Know vom Wall Street Journal .....	49
5.2.1 Allgemein .....	50
5.2.2 Analyse der Datenvisualisierung .....	51
5.2.3 Interaktionsmöglichkeiten .....	51
5.2.4 Bewertung.....	52
 <b>6 Schlussbetrachtung.....</b>	<b>53</b>
6.1 Zusammenfassung der Erkenntnisse .....	53
6.2 Ausblick auf den Datenjournalismus.....	54
6.3 Weiterführende Untersuchungen .....	55
 <b>Literaturverzeichnis .....</b>	<b>XI</b>
 <b>Eigenständigkeitserklärung .....</b>	<b>XXVI</b>

# Abkürzungsverzeichnis

App	=	Applikation
BBI	=	Berlin Brandenburg International
CC	=	Creative Commons
CSV	=	Comma-separated values
GPL	=	General Public License
HTML	=	Hypertext Markup Language
IT	=	Informationstechnik
KDD	=	Knowledge Discovery in Databases
MS	=	Microsoft
NDR	=	Norddeutscher Rundfunk
NYT	=	New York Times
OCR	=	Optical Character Recognition
PDF	=	Portable Document Format
RDF	=	Ressource Description Framework
Taz	=	die tageszeitung
USP	=	Unique Selling Proposition
WSJ	=	The Wall Street Journal
ZDF	=	Zweites Deutsches Fernsehen



# Abbildungsverzeichnis

Abbildung 1: Aufbau der Arbeit .....	3
Abbildung 2: Erwartetes weltweites Datenaufkommen bis 2015 .....	5
Abbildung 3: Rückgang der Auflagen und Anzahl von US-Tageszeitungen seit 1990 ..	9
Abbildung 4: Datenanalyse der US-Haushaltsprognosen seit 1980 .....	20
Abbildung 5: Spesenabrechnungen des damaligen Premierministers Gordon Brown..	22
Abbildung 6: Datenvisualisierung der Auswirkungen der Vorratsdatenspeicherung .....	24
Abbildung 7: Fluglärmkarte für den geplanten Flughafen BBI .....	25
Abbildung 8: Journalisten müssen lernen mit Spezialisten zusammenzuarbeiten .....	29
Abbildung 9: Erhöhung der Data-ink ratio eines Diagramms durch Reduktion .....	47
Abbildung 10: Beispiel einer Treemap von kategorisierten Nachrichten .....	48
Abbildung 11: Zusammenfassung der Farb-Grundsätze .....	49
Abbildung 12: Interaktive Datenvisualisierung „What They Know“ des WSJ .....	50

# Tabellenverzeichnis

Tabelle 1: Klassifikation der Techniken zur visuelle Datenanalyse.....	40
--	----

# 1 Einleitung

## 1.1 Problemhintergrund

Die Aufdeckung des Spesenskandals des britischen Parlaments durch den *Guardian* im Frühjahr 2009, der maßgeblich durch die mithilfe seiner Leser in der Analyse von hundertausenden Dokumenten ermöglicht wurde, verdeutlichte erstmals mit welchen Veränderungen der Journalismus in Zukunft konfrontiert werden wird:

In einer global vernetzten Welt entstehen mehr und mehr Stories durch Daten. Doch im Vergleich zu Augenzeugen, Regierungen und Pressestellen kommunizieren Daten ihre Geschichten nicht von alleine. Sie lagern in Archiven, Datenbanken und auf Websites, zwischen Texten, Tabellen und Quellcode. Journalisten und Medien, in ihrer Rolle als Berichterstatter, müssen lernen diese Geschichten in den Daten zu finden und der Öffentlichkeit sinnstiftend mitzuteilen. Darin besteht die Herausforderung für den Journalismus von morgen, dem Datenjournalismus.

Das interdisziplinäre Arbeitsfeld zwischen Investigativjournalismus, Statistik, Programmierung und Design, in dem sich Journalisten im Datenjournalismus bewegen, erweitert dabei die Anforderungen an ihren Berufsstand und erfordert eine Öffnung und Hinwendung der Medien zu ihrer Nutzerschaft, um erfolgreich mit Daten arbeiten zu können.

## 1.2 Zielsetzung

Mit dieser Arbeit soll gezeigt werden, welchen Einfluss Daten auf den Journalismus haben, welche Herausforderungen Medien dabei bestehen müssen und wie Daten den Journalismus verändern. Dabei geht die Arbeit auf das Potential des Datenjournalismus ein und zeigt zu welchen Zwecken und mit welchen Möglichkeiten Medien und Journalisten dieses Potential nutzen können.

Gleichzeitig soll die Arbeit beim Aufbau von Datenkompetenz helfen, die es Journalisten und Medien ermöglichen soll eigene Projekte im Datenjournalismus durchzuführen. Dazu werden die dafür notwendigen Techniken aufgezeigt, erklärt und in einen Workflow für datenjournalistische Projekte eingebettet.

Darüberhinaus soll diese Arbeit ein allgemeines Verständnis für den Umgang mit Daten im Journalismus aufbauen und für die zunehmende Dringlichkeit des Themas sensibilisieren.

Es wird erwartet, dass sich herausstellt, dass Daten für den Journalismus potentiell nützlich sind, aber die Erschließung dieses Nutzens mit wesentlichen Herausforderungen für Medien und Journalisten verknüpft ist.

## 1.3 Grenzen der Arbeit

Aufgrund seines interdisziplinären Charakters berührt Datenjournalismus sowohl technische als auch nicht technische Themenkomplexe. Dem Ansatz folgend, Datenjournalismus und dessen journalistisches Potential gesamtheitlich darzustellen, wird in dieser Arbeit zwar auf technologische Aspekte eingegangen, die Tiefe der Betrachtungen begrenzt sich aber auf eher allgemeine Erläuterungen und versucht zu technische Details unerwähnt zu lassen.

Die Auswertung sozialer Netzwerke durch *Opinion Mining* oder die Möglichkeiten der Semantik in der Strukturierung und Filterung von Daten konnten durch ihren thematischen und technologischen Umfang nicht in diese Arbeit aufgenommen werden und sind für die Untersuchung im Rahmen einer eigenständigen Arbeit geeignet.

Fast vollständig außer Acht lässt diese Arbeit den Einfluss datenjournalistischer Strukturen auf den Printjournalismus in Zeitungen und Zeitschriften. Die Begründung hierfür liefert der analoge und starre Charakter von Printmedien. Diese Einschränkungen des Mediums verhindern einen effektiven Einsatz datenjournalistischer Berichterstattungen, die von der Interaktion, Filterung und dynamischen Visualisierung von Daten leben.

## 1.4. Bedingungen der Arbeit

Während in den USA und Großbritannien vor allem einflussreiche Medien Datenkompetenz durch die Realisierung verschiedener Projekte aufbauen und bereits damit beginnen Datenjournalisten fest in ihre Redaktionen zu integrieren, spielt Datenjournalismus in deutschen Leitmedien kaum eine Rolle.

Daher standen mir bei der Bearbeitung des Themas vor allem englischsprachige Quellen zur Verfügung, die sich mit Datenjournalismus ganz oder zumindest in Ansätzen auseinandersetzen. Literatur, die sich konzentriert mit den Möglichkeiten eines datengestützten Journalismus auseinandersetzt, war jedoch zum Zeitpunkt der Entstehung dieser Arbeit nicht verfügbar.

Im Bereich der Datenanalyse, dessen Methoden das Gerüst für den in Kapitel 4 erarbeiteten Workflow bildeten, stützt sich meine Untersuchung hauptsächlich auf wissenschaftliche Einzelpublikationen oder Publikationen in Fachmagazinen.

Bei der Erarbeitung der Grundsätze von Datenvisualisierungen konnte ich zwar auf einige Literatur aus dem Informationsdesign zurückgreifen, diese ist im Bereich der Visualisierung umfangreicher und komplexer Datensätze jedoch nur auszugsweise zu gebrauchen und wird den aktuellen Anforderungen im Datenjournalismus nur teilweise gerecht.

## 1.5 Aufbau der Arbeit

### Einleitung

#### 2 Der digitale Wandel

Data Overload

Online vs. Print

Open Data

12

#### 3 Datenjournalismus

Datenpotential

Beispiele

Herausforderungen

15

#### 4 Datenworkflow

Vorbereitung

Aggregation

Aufbereitung

Datenanalyse

Präsentation

13

#### 5 Datenvisualisierung

Data-Ink

Lie Factor

Farbe

Schrift

Beispiel

8

### Schlussbetrachtung

Abbildung 1: Aufbau der Arbeit

**Kapitel 2:** Zu Beginn werden die Veränderungen der letzten Jahre in Technik, Medien und Gesellschaft aufgezeigt und deren Einfluss auf den Journalismus dargestellt.

**Kapitel 3:** Es wird das Potential von Daten für den Journalismus erläutert und aktuelle Beispiele aus dem Datenjournalismus gezeigt. Darauf aufbauend werden Herausforderungen für Medien und Journalisten im Datenjournalismus abgeleitet.

**Kapitel 4:** Als Leitfaden für Journalisten wird ein Workflow etabliert, der die erfolgreiche Realisierung datenjournalistischer Projekte von der Vorbereitung bis zur Veröffentlichung gewährleisten soll.

**Kapitel 5:** Eingängige und übersichtliche Datenvisualisierungen sind der Schlüssel bei der Präsentation der aus Daten gewonnenen Informationen. Daher werden Grundsätze guter Datenvisualisierungen vorgestellt und deren Wirkung an einem Beispiel geprüft.

## 1.6 Begriffserklärungen

### **Medien**

Der Begriff wird in dieser Arbeit vor allem in Bezug auf Online-Medien verwendet.

### **Leser/Nutzer**

Da diese Arbeit sich auf Online-Medien fokussiert, werden die Begriffe „Leser“ und „Nutzer“ synonym benutzt, für alle Personen die mediale Angebote jeglicher Art konsumieren.

### **Crowdsourcing**

Als Crowdsourcing wird die Auslagerung unternehmerischer Aufgaben durch den Einsatz der freiwilligen Arbeitskraft einer Vielzahl von Internetnutzern bezeichnet.

### **Stories/Geschichten**

Als Stories oder Geschichten in Daten werden in dieser Arbeit die Informationen und das Wissen bezeichnet, welches sich aus Daten gewinnen lässt und für eine journalistische Berichterstattung relevant ist.

### **Daten**

Sind Ansammlungen von Zeichen, die spezifische Informationen enthalten und elektronisch verarbeitet werden können.

### **Datensatz**

Ist eine Ansammlung gleich strukturierter Daten mit mindestens einem Datenfeld (Zelle).

### **Datenbank**

Ist eine Ansammlung von Datensätzen.

### **Dimension**

Eine Dimension ist ein Datensatz in einer Datenbank.

### **Algorithmus**

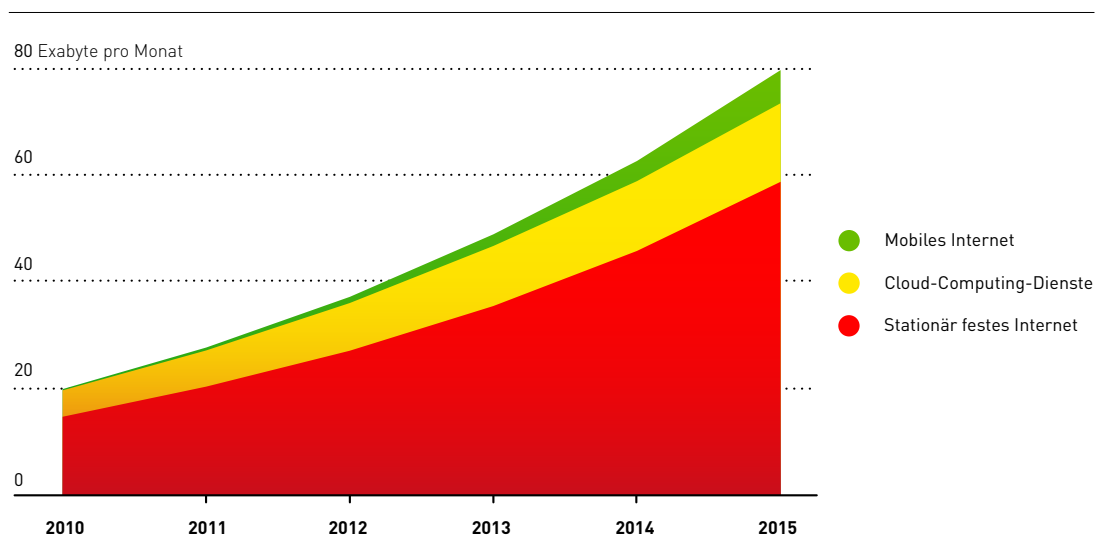
Ist eine nach Regeln definierte Handlungsvorschrift, mit der Daten nach spezifischen Mustern durchsucht und als Lösung ausgegeben werden.

### **Muster**

Sind Regelmäßigkeiten in Daten, die durch einen auf die Eingabedaten angewendeten Algorithmus ausgegeben werden. Durch die Analyse und Interpretation von Mustern kann neues Wissen aus Daten entstehen.

## 2 Der Digitale Wandel

Seit Johannes Gutenberg im 15. Jahrhundert den Buchdruck durch die Einführung beweglicher Lettern revolutionierte und den Siegeszug des Buches ausgelöst hatte, wächst die von der Menschheit produzierte Menge an Informationen stetig an. Durch die Digitalisierung wurde die Kommunikation mit den Jahrzehnten immer schneller, der Zugang zu Wissen immer einfacher und die Menge an zu bewältigenden Daten immer größer. Nach einer aktuellen Studie der *University of Southern California* wurden 2002 erstmals mehr digitale als analoge Daten gespeichert. Fünf Jahre später betrug der Anteil an digitalen Daten bereits 97 Prozent<sup>1</sup>. Und die Flut an Daten wächst stetig weiter. Nach dem *Cisco Visual Network Index 2011* wird sich das monatliche, weltweite Datenaufkommen von 20 Exabyte im Jahr 2010 auf 80 Exabyte bis 2015 vervierfacht haben (s. Abb. 2). Während sich das Wachstum des stationären Internets in dem Zeitraum kongruent zum allgemeinen Wachstum verhält, wird vor allem das Datenaufkommen des mobilen Internets stark zunehmen und 2015 das 26-fache seiner Größe von 2010 erreichen<sup>2</sup> und Internetnutzer werden Informationen vermehrt unterwegs konsumieren.



**Abbildung 2:**

Erwartetes weltweites Datenaufkommen bis 2015

**Datenquelle:**

Cisco Visual Network Index 2011

Doch trotz des einfachen Zugangs zum globalen Wissen durch das Internet ist das Leben im Informationszeitalter auch ein Kampf um die Informations- und Datenhoheit geworden. In diesem Kampf wird von Journalisten mehr und mehr Methodenkompetenz und Spezial-

1 Vgl. Wu (2011): How Much Information Is There in the World?, 18. Juni 2011

2 Vgl. Cisco Visual Networking Index (2011): White Paper, S. 2

wissen benötigt, mit dem sie die Herausforderungen des Informationszeitalters bestehen und Antworten im Informationschaos finden müssen. Online-Journalisten sind dabei im Besonderen gefordert.

Der hierfür verantwortliche Wandel im Umgang mit Daten und Informationen findet gleichzeitig auf technischer, medialer und gesellschaftlicher Ebene statt. Diese Ebenen werden in folgendem Kapitel aufgezeigt und erklärt, sowie ihr Einfluss auf den Online-Journalismus dargestellt. Denn um die Herausforderungen der Zukunft bestehen zu können, braucht es das Verständnis um die Entwicklungen der Vergangenheit und ihren Einfluss auf die Gegenwart. Im weiteren Verlauf der Arbeit ergeben sich daraus Handlungsanweisungen und Lösungsansätze, die einen effektiven Umgang mit Daten ermöglichen und den Anforderungen eines zukünftigen Online-Journalismus im Spannungsfeld des Informationszeitalters gerecht werden sollen.

## 2.1 Data Overload

*„As long as the centuries continue to unfold, the number of books will grow continually, and one can predict that a time will come when it will be almost as difficult to learn anything from books as from the direct study of the whole universe“*

(Denis Diderot, französischer Philosoph)<sup>3</sup>

Mit diesem Zitat des französischen Schriftstellers und Philosophen Denis Diderot aus dem Jahr 1755 kam erstmals die Idee einer permanenten Informationsüberflutung auf, die – wenn auch übertrieben – die Probleme der Gegenwart verdeutlicht. Die Kommunikation wurde globalisiert. E-Mails, Tweets, Artikel, Dokumente, Fotos und Videos sind heute in Sekunden über das Internet abrufbar, doch die Masse der verfügbaren Daten erschwert zunehmend die Suche nach den relevanten Informationen.

So hat jeder dritte Befragte einer repräsentativen Umfrage des Verbandes *BITKOM* das Gefühl, von Informationen überflutet zu werden<sup>4</sup>. Als Hauptverursacher wird von 71 Prozent der Teilnehmer das Fernsehen genannt, gefolgt vom Internet mit 43 Prozent und dem Radio sowie Printmedien mit jeweils 24 Prozent. Diese Einschätzung ist besonders für das Internet interessant, ist es doch von den Aufgezählten das einzig aktive Medium, in welchem Nutzer ihre gewünschten Inhalte selbst auswählen können und welches über Möglichkeiten zur Filterung von Informationen verfügt. Allerdings ist das Internet auch das einzige aktive Medium, in dem Nutzer Inhalte selbst und zu jeder Zeit erstellen können.

Dieser User Generated Content ist neben der steigenden Verbreitung des Internets<sup>5</sup> ein maßgeblicher Grund für den Anstieg der Datenmengen in den letzten Jahren<sup>6</sup>. Durch die Verbreitung partizipativer Technologien und die Etablierung sozialer Netzwerke seit dem Jahr 2000 wurde das World Wide Web zum Social Web, in dem Nutzer schnell und einfach selbst Websites aufbauen, miteinander interagieren und eigen- oder fremdgenerierte Daten

3 Vgl. Rosenberg (2003): Journal of the History of Ideas, S. 1, zit. nach Diderot (1755)

4 Vgl. BITKOM (2011): Information Overload?, S. 7

5 Vgl. o.V. (2010): Über zwei Milliarden Internet-Nutzer, 18. Juni 2011

6 Vgl. z.B. Twitter (2011): 200 million Tweets per day; Facebook (o.J), Statistics, 18. Juli 2011



kommunizieren können. Große soziale Netzwerke wie *Facebook* oder *Twitter* verursachen mit ihren 750 Millionen<sup>7</sup> und 200 Millionen Mitgliedern<sup>8</sup> täglich eine große Datenmasse, die die Netzwerke zu den trafficstärksten Websites der Welt hat wachsen lassen<sup>9</sup>.

Einen weiteren Grund der Informationszunahme und Datenüberflutung sieht der US-amerikanische Autor und Internetberater Clay Sharky im ökonomischen Wandel, den das Internet verursacht hat. Seit Gutenberg war das Geschäftsmodell von Medien immer gleich: Medien produzieren zu Beginn ihrer Wertschöpfung auf eigene Kosten Bücher, Zeitungen, Radio- und Fernsehsendungen und refinanzieren sich erst im Nachhinein über die Einnahmen aus dem Verkauf ihrer Produkten und der darin geschalteten Werbung. Durch dieses finanzielle Risiko waren sie gezwungen ihre Inhalte nach qualitativen Kriterien zu filtern und nur die Inhalte zu veröffentlichen, die ihre Zielgruppe ansprechen<sup>10</sup>. Das änderte sich mit dem Internet, als die Veröffentlichung von Inhalten nicht mehr an Produktionsgüter, technische Infrastrukturen, oder Produktionsstätten gebunden war. Ein Computer mit Strom- und Internetanschluss reichte fortan aus, wodurch das Risiko verschwand und eine „post-Gutenbergsche Ökonomie“ etablierte, in der jeder Nutzer seine Inhalte kostenlos publizieren kann<sup>11</sup>. Gleichzeitig verschwand mit dem Risiko aber auch der Ansatz einer qualitativen Filterung, weswegen heute mehr Content mit sinkender Relevanz entsteht, so Sharky<sup>12</sup>.

Trotz dieser Entwicklungen fühlten sich über die Hälfte der Befragten der BITKOM-Studie zwischen 14 und 29 noch nie von Informationen überflutet<sup>13</sup>. Ein Grund hierfür ist, dass die meisten mit diesen Medien aufgewachsen sind und bezüglich des Internets bereits in der Schule, im Studium oder später im Beruf Methoden erlernt haben, wie sich mit welchen Werkzeugen Informationen finden und filtern sowie Daten strukturieren lassen. Da die Idee der Informationsüberflutung bereits 1755 durch Denis Diderot und der Begriff 1970 durch den US-amerikanischen Schriftsteller und Futurologen Alvin Toffler eingeführt wurden<sup>14</sup>, ist das Problem also keineswegs auf das World Wide Web zurückzuführen. Konkret besteht eine Informationsüberflutung spätestens seit dem Bestehen der Bibliothek von Alexandria zwischen dem 3. und 1. Jh. v. Chr., als zum ersten Mal in der Geschichte der Menschheit mehr Informationen verfügbar waren, als ein Mensch in seinem Leben verarbeiten kann<sup>15</sup>. So ein immer gleiches Problem, was über eine lange Zeit besteht, ist nach Sharky kein Problem, sondern ein Fakt<sup>16</sup>. Problematisch ist also nicht die Informationsüberflutung selbst, sondern vielmehr der Mangel an Erfahrungen im Umgang mit komplexen Informationen und somit ein Mangel in deren Filterung, der Nutzern im Umfeld von Daten die Orientierung raubt.

Darunter fällt auch ein Mangel im Umgang mit Filter-Werkzeugen, der auch dadurch entstehen kann, dass geeignete Werkzeuge noch nicht existieren oder nahezu unbekannt sind. War eine Literaturrecherche in den 1980er Jahren noch hauptsächlich „Handarbeit“,

7 Vgl. Facebook (o.J.): Statistics, 18. Juli 2011

8 Vgl. Kohlhepp (2011): Twitter feiert fünften Geburtstag, 18. Juli 2011

9 Vgl. Alexa Top Sites (o.J.): <http://bit.ly/alexatop>, 18. Juni 2011

10 Vgl. Sharky (2008): <http://bit.ly/claysharky> (Video, ab 3:00 bis 5:00)

11 Vgl. ebenda

12 Vgl. ebenda

13 Vgl. BITKOM-Studie Information Overload? (2011): „Die jüngere Generation hat damit kaum Probleme“

14 Vgl. Toffler (1970): Future Shock, S. 350

15 Vgl. Juskalian (2008): Interview with Clay Shirky, Part I, 18. Juni 2011

16 Vgl. Sharky (2008): <http://bit.ly/claysharky> (Video, ab 22:00 bis 23:40)

bei der jedes Buch gelesen und von selbst auf die nötige Relevanz geprüft werden musste, ermöglichen heute Werkzeuge wie *Google Books*, *Amazons* Nutzerempfehlungen und offene Bibliotheksnetzwerke mit ihren umfangreichen Datenbanken in wenigen Minuten eine Vielzahl von relevanten Quellen aus einer großen Datenmasse herauszufiltern, die ohne den Einsatz solcher Technologien undurchschaubar bleiben würden. Daten helfen demnach mehr Relevanz und Wissen zu generieren, als sie die Menschen mit ihren Informationen überfluten – vorausgesetzt Nutzer besitzen die Kompetenz mithilfe von Werkzeugen mit Daten umzugehen.

Hier zeigt sich die Bedeutung von Arbeitstechniken und Werkzeugen bei der Arbeit mit Daten, die in Kapitel 4 genauer vorgestellt werden und die Journalisten benötigen, wollen sie ihrer Rolle im Informationszeitalter gerecht werden.

## 2.2 Internet Killed the Newspaper Star

Als Mitte der 1990er Jahren die ersten Verlage mit Nachrichtenwebsites online gingen<sup>17</sup>, um das Internet als zusätzlichen Kommunikationskanal zu nutzen und sich ein zweites Standbein aufzubauen, begingen sie nach dem US-amerikanischen Journalisten und Schriftsteller Tony Rogers einen für ihre Industrie folgenschweren Fehler: Sie stellten ihren Inhalt – anfangs noch Kopien aus den Printprodukten – kostenlos im Internet bereit<sup>18</sup>. Mit Beginn des neuen Jahrtausends erkannte die steigende Zahl der Internetnutzer<sup>19</sup>, dass der gleiche Inhalt, für den sie bei den Printausgaben bezahlen mussten, online auch gratis zur Verfügung steht und sie wanderten mehr und mehr ins Netz ab<sup>20</sup>, wo sich die Gratiskultur als Norm aller online verfügbaren Inhalte etablierte<sup>21</sup>.

Gleichzeitig führten Rückgänge im Anzeigengeschäft in den Jahren 2000 bis 2002 zu enormen wirtschaftlichen Verlusten bei Verlagen<sup>22</sup>, die vor allem durch den schrumpfenden Anteil der Stellenanzeigen im Printprodukten zu Gunsten von Jobportalen im Internet<sup>23</sup> und die sinkenden Werbeausgaben der IT-Branche<sup>24</sup> nach der Dotcom-Blase von 2000 verursacht wurden.

Mit steigenden Nutzerzahlen, eigenen Angeboten und erhöhter Aktualität entwickelten sich die Websites von Zeitungsverlagen in den folgenden Jahren zu eigenständigen Nachrichtenportalen weiter, die vor allem mit kostenlos verfügbaren, überregionalen Inhalten, Druck auf überregionale Printtitel ausübten<sup>25</sup>. Seit 2005 können deutsche Nachrichtensites als direkte Konkurrenz von Zeitungen betrachtet werden, da die Zahl der täglichen Intensivnutzer überregionaler Nachrichten im Internet in diesem Jahr gleich auf mit der Zahl der Intensivleser von Zeitungen lag<sup>26</sup>. In den USA verliefen die

17 Vgl. z.B. o.V. (2000): Aus STERN Online wird künftig die Marke stern.de; Keese (2007), Willkommen bei WELT ONLINE; 18. Juli 2011

18 Vgl. Rogers (o.J.): Why Are Newspapers Dying?, 18. Juni 2011

19 Vgl. Deutsches Institut für Wirtschaftsforschung e.V. (2003), Internetnutzung in Deutschland, 18. Juni 2011

20 Vgl. o.V. (2007): Studie: Tageszeitungsleser wandern ab ins Internet, 18. Juni 2011

21 Vgl. Darnstädt (2007): Die Gratis-Kultur, 18. Juli 2011

22 Vgl. Meyer-Lucht (2003): Sinkende Auflagen, Einbrüche im Anzeigengeschäft, Konkurrent Internet; S. 6

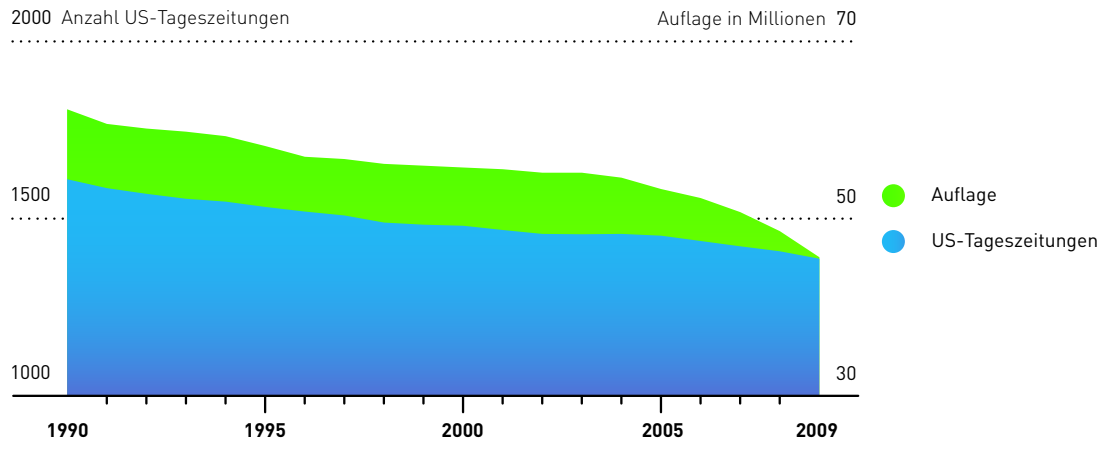
23 Vgl. Schmitz (2003): Online-Stellenbörsen profitieren von Anzeigenrückgang bei Zeitungen, 18. Juli 2011

24 Vgl. Meyer-Lucht (2003): Sinkende Auflagen, Einbrüche im Anzeigengeschäft, Konkurrent Internet; S. 7

25 Vgl. ebenda, S. 5

26 Vgl. Kolo/Meyer-Lucht (2007): Erosion der Intensivleserschaft, S. 527

Entwicklungen ähnlich. Dort konnte schon ab 2004 ein deutlich zunehmender Auflagenrückgang beobachtet werden (s. Abb. 3).



**Abbildung 3:**

Rückgang der Auflagen und Anzahl von  
US-Tageszeitungen seit 1990

**Datenquelle:**

Newspaper Association of America,  
Total Paid Circulation

Trotz des wachsenden Konsums von Nachrichten im Internet waren Nachrichtensites gegenüber Zeitungen unrentabler, da „Verlage [...] weltweit Milliarden in die neue Technik [investierten], ohne je angemessene Erträge einzufahren“<sup>27</sup>.

Deshalb suchten die Verlage ab Mitte der 2000er Jahre nach neuen Möglichkeiten ihre Online-Inhalte stärker zu monetarisieren, um damit auch die Verluste des Printgeschäfts auszugleichen. So führte die *New York Times* 2005 mit *TimesSelect* ein durchaus erfolgreiches Bezahlssystem auf ihrer Website ein, allerdings war das Abonnenten-Modell nur temporär profitabel<sup>28</sup>. Denn zunehmend bestimmten Suchmaschinen, die auf die Verfügbarkeit freier Daten angewiesen sind, den Großteil des Datenverkehrs auf Nachrichtenwebsites<sup>29</sup>. Vor allem *Google* und dessen Nachrichten-Aggregator *Google News* halfen den Nutzern mit ihren Möglichkeiten zur intelligenten Filterung eine größere Menge relevanter Inhalte in kürzerer Zeit zu finden. Durch die Sperrung der Inhalte hinter der mit *TimesSelect* eingeführten Bezahlschranke schloss sich die *New York Times* von diesem Nutzerkreis aus. Da der dadurch verursachte Reichweiteinbruch zu größeren Verlusten im wachsenden Online-Anzeigenmarkt führte, als die Times durch die Einnahmen aus dem Abo-Geschäft generieren konnte<sup>30</sup>, wurde *TimesSelect* genau zwei Jahre nach dessen Einführung wieder eingestellt<sup>31</sup>. Die Auflösung des Abo-Modells der *New York Times* führte auch zu einem vorläufigen Strategiewechsel der Zeitung, die nicht nur wieder ihre Online-Inhalte jedem Nutzer frei zur Verfügung stellte, sondern auch ihre Archive ab 1987 für Suchmaschinen öffnete<sup>32</sup>,

27 Eberle et al. (2010): Das Ende der Gratiskultur im Internet ist gekommen, S. 1

28 Vgl. Pérez-Peña (2007): Times to Stop Charging for Parts of Its Web Site, 18. Juni 2011

29 Vgl. ebenda

30 Vgl. Kramer (2007): New York Times To Close TimesSelect Effective Midnight Tuesday, 18. Juni 2011

31 Vgl. Schiller (o.J.): A Letter to Readers About TimesSelect, 18. Juni 2011

32 Vgl. ebenda

um über den zusätzlich online verfügbaren Inhalt ihre Reichweite und Werbeeinnahmen ausbauen zu können<sup>33</sup>. Ende März 2011 hat die *New York Times* wieder eine Paywall installiert<sup>34</sup>. Ob diese sich diesmal für die Zeitung dauerhaft rechnet, bleibt abzuwarten.

Seit dem Jahr 2005 nutzt die Öffentlichkeit das Internet durch die Verbreitung von Strukturen des *Social Web* vermehrt als aktives Medium. Privatnutzer, die sich verstärkt eigene Websites und Blogs erstellten, berichten online über ihr Leben, ihr Umfeld und aktuelle Geschehnisse<sup>35</sup> und begründeten als Bürgerjournalisten einen neuen Medienkanal. Aufgrund fehlender Messungen kann hier keine Aussage zur Gesamtzahl der vorhandenen Blogs getroffen werden, aber die Mitgliederzahlen der beliebten Blogging-Plattformen *Wordpress* und *tumblr*, die mit 50 Millionen<sup>36</sup> und 21 Millionen<sup>37</sup> angemeldeten Mitgliedern zu den größten ihre Art gehören, skizzieren das Ausmaß der Blogosphäre und der potentiellen Meinungsmacht der Öffentlichkeit gegenüber den etablierten Medien.

Neben dieser Zunahme privater Meinungen im Netz gründeten sich ab Mitte der 2000er Jahre auch immer mehr Fachblogs<sup>38</sup>, die sich mit spezifischen und oft sehr speziellen Themen auseinandersetzten und mit steigender Reichweite zu eigenständigen Onlinemagazinen und selbst angesehenen Medien heranwuchsen. Bekannte Beispiele hierfür sind der Technik-Blog *Engadget* und der liberale Politik- und Nachrichten-Blog *The Huffington Post*. Auf der einen Seite entstand so mit der stärker kommunizierenden Öffentlichkeit und ihren vielschichtigen, oft persönlichen Themen ein Gegenpol zum Informationsangebot der Leitmedien, deren Inhalte durch Nachrichtenagenturen und die selektiven Filterprozesse von Journalisten bestimmt werden<sup>39</sup>.

Auf der anderen Seite entwickelte sich durch das starke Wachstum von Netzwerken wie *Facebook* und *Twitter* ein neuer, reichweitenstarker Distributionskanal für die Leitmedien, indem deren Inhalte neben denen der Masse von den Nutzern selbst in den sozialen Medien geteilt wurden. Hier entstand ein Vorteil für die etablierten Medien, vermischten sich doch Inhalte von Fachblogs, Bürgerjournalisten und Privatpersonen miteinander und kommunizierten fortan direkt auf einer Ebene. Noch nie zuvor in der Geschichte waren Journalisten mit ihren Inhalten näher bei ihren Lesern als heute und noch nie waren die Menschen gleichzeitig mit ihren Geschichten so nah an den Leitmedien.

Journalisten können daher Social Media-Kanäle sowohl als Quelle interessanter Inhalte und neuer Themen nutzen, als auch zum Aufbau oder zur Anreicherung ihrer journalistischen Beiträge mit Stimmungen und Meinungen aus der Öffentlichkeit. Besonders hilfreich kann der Einsatz von Social Media-Inhalten dabei in der Berichterstattung über Krisenregionen sein. So stützte sich ein Großteil der westlichen Berichterstattung über die im Dezember 2010 begonnenen Proteste in der arabischen Welt auf Videos, Fotos und Textnachrichten aus sozialen Netzwerken, die die Welt noch über das Geschehen informieren konnten, als die meisten Journalisten die Krisenregionen schon verlassen hatten<sup>40</sup>.

---

33 Vgl. Pérez-Peña (2007): Times to Stop Charging for Parts of Its Web Site, 18. Juli 2011

34 Vgl. Lischka (2011): „New York Times“ führt das Digitalabo ein, 19. Juni 2011

35 Vgl. Schmidt (2005): Der Weg zum eigenen Blog, 18. Juli 2011

36 Vgl. Klaß (2011): Heimat von über 50 Millionen Blogs, 18. Juli 2011

37 Vgl. Lipowicz (2011): More federal agencies blog on Tumblr, 18. Juli 2011

38 Vgl. z.B. Engadget (o.J.): About Us; o.V. (2005), Wo Promis endlich reden dürfen, 18. Juli 2011

39 Vgl. Borchardt (2011): Brecht das Informationsmonopol!, 18. Juli 2011

40 Vgl. Weaver/Hill (2011): Yemen, Syria and Middle East unrest, 19. Juni 2011

Als die Leitmedien das Potential von *User Generated Content* erkannten<sup>41</sup>, soziale Netzwerke für die Arbeit einsetzten, vermehrt Inhalte aus Blogs übernahmen und ihre Nachrichtenportale mit Kommentarfunktionen, eigenen Blogs, Umfragen und Foren ausstatteten, ermöglichten sie gleichzeitig der Öffentlichkeit ihre Meinung breiter in ihre Medienkanäle zu tragen. Die Medien hatten sich aneinander angenähert und demokratisiert<sup>42</sup>. Die Leserschaft konnte nun mitentscheiden, was wichtig ist und Themen auf die Nachrichtenagenda setzen, was vormals nur Nachrichtenwebsites, Zeitungen, Radio- und Fernsehsendern vorbehalten war<sup>43</sup>.

Ein Beispiel hierfür ist der Ortungsdaten-Skandal von *Apple*. Im April 2011 veröffentlichten der *iPhone*-Entwickler Alasdair Allan und der Informatiker Pete Warden ihre Entdeckung, dass *Apples* mobiles Betriebssystem *iOS* hintergründig ein Bewegungsprofil seiner Nutzer erstellt<sup>44</sup>. Das Thema beherrschte schnell die Medien und selbst Verbraucherschutzministerin Ilse Aigner schaltete sich ein<sup>45</sup>. Der Skandal endete in einem Update des Betriebssystems durch *Apple*, das die Speicherung der Daten auf eine Woche beschränkt<sup>46</sup>.

Durch das Internet stehen den etablierten Medien heute „tausende Experten“ und „ein wahrer Schatz an Informationen“ zur Verfügung, wie der Chefredakteur des Guardian, Alan Rusbridger, die Entwicklung beschreibt<sup>47</sup>. Diese Informationen im Zusammenspiel mit den sozialen Medien zu finden, zu nutzen und durch das Wissen der Masse für eine ausgewogenere, aktuellere und unabhängigere Berichterstattung zu sorgen, wird eine der kommenden Herausforderungen für den Online-Journalismus werden. Keinesfalls wird dieser durch soziale Medien überflüssig. Mit ihrer enormen Reichweite und Vielfalt werden Nachrichtenportale auch weiterhin die Masse erreichen. Doch diese ist besser informiert und kritischer geworden. Folglich müssen die Medien auch lernen, die Informationen, die ihnen zur Verfügung stehen, besser nach Relevanz zu filtern und die Stories in den Daten finden. Gleichzeitig müssen sie sich kompetenter positionieren und dort berichten, wo der Masse das nötige Wissen fehlt, um Zusammenhänge und Entwicklungen zu erkennen. Neben einer größeren Fachkompetenz brauchen Online-Journalisten von morgen also auch Methodenkompetenz im Umgang mit Daten.

## 2.3 Open Data

Die Menschen, die am 10. Juni 2011 in Dresden öffentlich gegen die einen Tag vorher vorgenommene Schließung der Filmpiraterie-Website *kino.to* demonstrierten<sup>48</sup>, sind bezeichnend für die Ansichten und Denkweise der vernetzten Gesellschaft gegenüber Medieninhalten. Die Gratiskultur ist zu einem Bestandteil des öffentlichen Lebens geworden, der Konsum von kostenlosen Inhalten zu einer Selbstverständlichkeit, die Forderung

41 Vgl. z.B. Ostrow (2008): iReport Goes Live; CNN Gets Citizen Journalism; Conlan (2006): BBC to pay for viewers' pictures, 18. Juli 2011

42 Vgl. Meckel (2006): Was vom Tage übrig bleibt, 19. Juni 2011

43 Vgl. O'Reilly (2005): What is Web 2.0, 19. Juni 2011

44 Vgl. Alan/Warden (2011): Got an iPhone or 3G iPad? Apple is recording your moves, 18. Juli 2011

45 Vgl. o.V. (2011): Aigner fordert Klarheit von Apple, 19. Juni 2011

46 Vgl. o.V. (2011): Update: iPhone soll jetzt weniger Daten sammeln, 18. Juli 2011

47 Grimberg (2009): „Das alte System ist kaputt“, 19. Juni 2011

48 Vgl. Noack (2011): „Wir alle schauen kino.to“, 19. Juni 2011



nach mehr Transparenz, Offenheit und Freiheit ein Merkmal des Informationszeitalters. Die Befürworter einer offenen Gesellschaftsform finden sich dabei in der Open-Data-Bewegung vereint. Hinter der Grundidee, dass die Abschaffung von „*Datenschutz-, Sicherheits- oder Zugangsbeschränkungen*“<sup>49</sup> und die Herausgabe von Inhalten, Wissen und Daten zur freien Verwendung unter anderem in Kultur, Wissenschaft und Politik zu positiven Entwicklungen für die Gesellschaft führt<sup>50</sup>, entstanden verschiedene „Open“-Initiativen.

### 2.3.1 Open Content und die Freiheit der Inhalte

Inspiziert durch die von dem Programmierer Richard Stallmann 1989 veröffentlichte *GPL-Lizenz* zur Lizenzierung freier Software und dem dadurch einsetzenden Erfolg von Open-Source-Programmen wie *Linux*, entwickelte sich parallel dazu der Open-Content-Gedanke, geprägt vom US-amerikanischen Lernpsychologen David Wiley<sup>51</sup>. Open Content steht für den freien Zugang, die freie Anwendung, uneingeschränkte Modifikation und Weiterverbreitung von Texten, Bild- und Tonwerken bei bestehendem urheberrechtlichem Schutz durch freie Lizenzen<sup>52</sup>. Der Gedanke gewann 2001 mit der Einführung der *Creative Commons-Lizenz* (CC-Lizenz) durch den Rechtsprofessor Lawrence Lessig an Popularität. Denn erstmals ermöglichte *Creative Commons* ein für Urheber flexibles Modell zur Gestaltung ihrer in Werken freigegebenen Rechte und stellte Lizenzen<sup>53</sup> bereit, „*die bei der Verbreitung kreativer Inhalte genutzt werden können, um die rechtlichen Bedingungen festzulegen*“<sup>54</sup>.

Neben *Flickr*, dem größten sozialen Bildernetzwerk, dass die Nutzung von *CC-Lizenzen* für hochgeladenes Material ermöglicht<sup>55</sup>, nutzt auch die freie Enzyklopädie *Wikipedia* seit 2009 eine Form der Lizenz für ihre Inhalte<sup>56</sup>. *CC-Lizenzen* finden auch in einigen Medien bereits Verwendung. So zum Beispiel auf dem IT-Nachrichtenportal *netzwelt.de*, dass seine Artikel und Nachrichten unter einer *CC-Lizenz* zur freien Nutzung bereitstellt<sup>57</sup>, oder in dem von Mario Sixtus betriebenen Video-Podcast *Elektrischer Reporter*, der auch im *ZDF* ausgestrahlt wird und seine Videos ebenfalls unter einer *CC-Lizenz* veröffentlicht<sup>58</sup>. Früh engagierte sich auch der *NDR* bei der Veröffentlichung freier Inhalte, der seine Medienmagazine *Zapp* und *Extra3* schon seit 2007 online unter einer Creative Commons-Lizenz anbietet<sup>59</sup>. Diese Beispiele belegen ein steigendes Bewusstsein für die „Befreiung“ von Inhalten aus starren Lizenz- und Urheberrechtsmodellen, nicht nur in den Forderungen der Open-Content-Bewegung, sondern auch bereits in den Köpfen deutscher Medienmacher.

49 Open Data Network (2010): Open Government Data Principles, 18. Juli 2011

50 Vgl. Weiss (2002): Borders in Cyberspace: Conflicting Public Sector Information Policies and their Economic Impacts, S. 1

51 Vgl. Grossmann (1998): New Free License to Cover Content Online, 18. Juli 2011

52 Vgl. The Free Content and Expression Definition (o.J.): Definition, 19. Juni 2011

53 Vgl. Creative Commons (o.J.): About The Licenses, 19. Juni 2011

54 Creative Commons (o.J.): Was ist CC?, 19. Juni 2011

55 Vgl. Flickr (o.J.): How can I copyright my photos?, 19. Juni 2011

56 Vgl. o.V. (2009): Wikipedia Moving From GFDL To Creative Commons License, 19. Juni 2011

57 Vgl. H2 media factory (o.J.): Rechtliche Hinweise, 18. Juli 2011

58 Vgl. Sixtus (2010): Labor-Wochen beim Elektrischen Reporter, 18. Juli 2011

59 Vgl. Voß (2007): „Zapp“ und „Extra 3“ unter Creative-Commons-Lizenz, 18. Juli 2011

### 2.3.2 Open Access und die Freiheit der Wissenschaft

Obwohl der Begriff Open Data relativ neu ist, ist der Gedanke dahinter schon über ein halbes Jahrhundert alt: So wurden während dem *Internationalen Geophysikalischen Jahr* von 1957 bis 1958 Welt-Datenzentren eingerichtet, um die aus der wissenschaftlichen Arbeit gewonnenen Daten „vor einem katastrophalen Verlust zu schützen“<sup>60</sup>, zu bündeln und deren Zugang für die Wissenschaftler zu vereinfachen<sup>61</sup>. Der Zugang zu diesen Daten war schon damals frei und es fielen lediglich Kosten zum Kopieren und Übersenden des Materials an<sup>62</sup>.

Mit dem Aufkommen des Internets und der zunehmenden Digitalisierung war es technisch nun möglich, einen komplett kostenfreien Zugang zum globalen Wissen aufzubauen. So firmierte sich zu Beginn der 1990er Jahre die internationale Open-Access-Bewegung. Begünstigt wurde diese Entwicklung durch die *Zeitschriftenkrise* Mitte der 1990er Jahre, in der durch die Verteuerung von wissenschaftlichen Zeitschriften viele Universitätsbibliotheken ihre Bestände einkürzen mussten und so Wissenschaftlern der Zugang zu aktuellen Wissensquellen fehlte<sup>63</sup>. Hinzu kamen die harten Bedingungen, unter denen Wissenschaftler ihre Arbeitsergebnisse unter Verlagen veröffentlichen mussten<sup>64</sup>. Ziel der Open-Access-Bewegung ist die Einrichtung von freien, öffentlich zugänglichen Dokumentenarchiven und Bibliotheken die wissenschaftliche Literatur ohne „finanzielle, gesetzliche oder technische Barrieren“<sup>65</sup> beinhaltet. Begründet wird diese Forderung durch den Sachverhalt, dass „wissenschaftliche Informationen als in der Regel durch öffentliche Mittel subventionierte Ergebnisse der Wissensproduktion und daher als Gemeinschaftsgut [...] für alle Interessierten ohne Nutzungsentgelte zugänglich sein sollten“<sup>66</sup>.

Einer der Meilensteine der Bewegung war die Gründung der Budapest Open-Access-Initiative im Dezember 2001, in deren Grundsatzerklärung Möglichkeiten zu einer erfolgreicherer, globalen Zusammenarbeit und Strategien zur Wahrung der Interessen von Wissenschaftlern und Institutionen festgehalten wurden<sup>67</sup>. Ein weiterer Meilenstein war die *Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen* von 2003, in der sich deutsche und internationale Forschungsinstitutionen dazu bekannten, „die Weiterentwicklung des Open-Access-Gedankens zu unterstützen“ und „Forscher darin [zu] bestärken, ihre Ergebnisse Open Access zu veröffentlichen“<sup>68</sup>.

Eine für Journalisten spannende Entwicklung findet sich im *Linked-Open-Data-Projekt* des *Word Wide Web Consortiums*. Ziel des Projekts ist es, mithilfe semantischer Technologien alle im Web frei zugänglichen Informationen und Datenbestände miteinander zu verknüpfen<sup>69</sup> und Informationen über Daten und *Metadaten* über die Grenzen von Websites, Domänen und Servern bereitzustellen, damit durch die restriktionsfreie Weiterverwendung der Daten Verbindungen und Zusammenhänge zwischen ihnen deutlich gemacht werden

60 World Data Center System (o.J.): About the World Data Center System, 19. Juni 2011

61 Vgl. World Data Center System (o.J.): The World Data Center System, 18. Juli 2011

62 Vgl. ebenda

63 Vgl. open-access.net (o.J.): Die Geschichte der Open-Access-Bewegung, 19. Juni 2011

64 Vgl. Dambeck (2004): Aufstand gegen die Hüter des Wissens, 18. Juli 2011

65 open-access.net (o.J.): Was ist Open Access, 19. Juni 2011

66 Mruck et al. (2004): Open Access: (Social) Sciences as Public Good, S. 1

67 Vgl. Budapest Open Access Initiative (o.J.): <http://bit.ly/budapestopen>, 18. Juli 2011

68 Max-Planck-Gesellschaft (o.J.): Berliner Erklärung, 19. Juni 2011

69 Vgl. Mühlberger (o.J.): Linking Open Data: Anwendungen für Endanwender, S. 1

können<sup>70</sup>. Außerdem soll dadurch der Aufbau eines semantischen Webs vorangetrieben werden<sup>71</sup>. Die für die Verknüpfung obligatorischen Metadaten werden über *RDF-Links* beschrieben, die Verbindungen zwischen den Daten herstellen und diese so erst maschinenlesbar machen. Nach dem Erfinder des World Wide Web Tim Berners-Lee ist „*RDF [...] to data what HTML is to documents*“<sup>72</sup> und hilft Zusammenhänge zwischen Daten durch „triples“ herzustellen, die einen Bezug zwischen einem „*subject*“ und einem „*object*“ über dessen Eigenschaft (*predicate*) herstellen<sup>73</sup>.

Die durch die Vernetzung der Daten entstehende Struktur ermöglicht semantischen Suchmaschinen komplexe Anfragen auf Grundlage der Daten und mithilfe von Filtersystemen zu lösen. Dadurch können zum Beispiel bei einer Recherche über Hochhäuser in Hong Kong mit mehr als 50 Stockwerken exakt diese Kriterien in eine semantische Suchmaschine eingegeben werden, die dann nach passenden Ergebnissen in einer Gebäude-Datenbank sucht<sup>74</sup>. Die Suche nach solch präzisen Kriterien kann Journalisten dabei helfen wertvolle Zeit bei der Rechercharbeit zu sparen und die Qualität der Ergebnisse ihrer Arbeit zu verbessern.

### 2.3.3 Open Government und die Freiheit öffentlicher Daten

Staaten sammeln ständig Daten, um über die laufenden Veränderungen von Wirtschaft, Bevölkerung und Politik informiert zu bleiben. Allein in Deutschland gibt es mit den statistischen Landesämtern und dem statistischen Bundesamt mehrere Behörden, die zum Teil im gleichen Gebiet aktiv sind. Obwohl viele der gesammelten öffentlichen Daten bereits online verfügbar sind, behindern *proprietäre Dateiformate*, fehlende Rohdaten, undurchsichtige Copyright-Bestimmungen und eingeschränkte Nutzungsbedingungen<sup>75</sup> eine effektive Weiterverarbeitung der Daten und somit den Einsatz für eine breite Öffentlichkeit. Glücklicherweise beginnt in einigen Ländern bereits die Auflösung dieser Restriktionen durch die Verbreitung der Idee der Open-Government-Bewegung. Diese fordert den Zugang zu im Web miteinander vernetzten Daten von Staat und Verwaltung „*im Interesse der Allgemeinheit ohne jedwede Einschränkung zur freien Nutzung, zur Weiterverarbeitung und zur freien Weiterverwendung*“<sup>76</sup>.

Als der frühere Premierminister Großbritanniens Gordon Brown den Erfinder des World Wide Web, Tim Berners-Lee, im Frühjahr 2009 fragte, wie sein Land das Internet besser nutzen kann, antwortete Berners-Lee, die Regierung solle all ihre Daten online verfügbar machen<sup>77</sup>. Das Ergebnis dieses Gesprächs war der Launch der Website *data.gov.uk* im Januar 2010<sup>78</sup>, die eine Vielzahl von Regierungsdaten online und in maschinenlesbarer Form der Öffentlichkeit

70 Vgl. von Lucke/Geiger (2010): Open Government Data: Frei verfügbare Daten des öffentlichen Sektors, S. 3

71 Vgl. Linked Open Data (o.J.): Why do you think this project will have a wide impact?, 19. Juni 2011

72 Berners-Lee (2007): Hearing on the „Digital Future of the United States: Part I -- The Future of the World Wide Web“, S. 5

73 Vgl. Bizer et al. (o.J.): How to Publish Linked Data on the Web, The RDF Data Model, 18. Juli 2011

74 Vgl. DBpedia, <http://bit.ly/rdfwolkenkratzer>

75 Vgl. z.B. Statistisches Bundesamt Deutschland (o.J.): Hinweise zum Datenschutz;  
Eurostat (o.J.): Rechtlicher Hinweis; Gegenteil:  
Data.gov (o.J.): Data Policy, 9. Juni 2011

76 von Lucke/Geiger (2010): Open Government Data: Frei verfügbare Daten des öffentlichen Sektors, S. 3

77 Vgl. Rogers (2010): Information is power, 19. Juni 2011

78 Vgl. data.gov.uk (o.J.): What is our approach?, 18. Juli 2011



frei zugänglich macht. Inspiriert wurde das Portal dabei von seinem amerikanischen Pendant, *Data.gov*, das von der Obama-Regierung bereits im Mai 2009 gestartet wurde<sup>79</sup>. Damit sicherte sich Großbritannien im Bereich Open Government die Vorreiterrolle in Europa, denn außerhalb des Vereinigten Königreichs gibt es kein vergleichbares Projekt.

In seinem 2002 veröffentlichten Bericht *Borders of Cyberspace: Conflicting Public Sector Information Policies and their Economic Impact* führt der ehemalige Mitarbeiter des National Weather Service Peter Weiss die Sichtweise vieler europäischen Behörden als Hauptgrund für das Fehlen ähnlicher Plattformen auf. Nach Weiss sehen die Behörden ihre mit öffentlichen Geldern finanzierten Daten als handelbare Ware zur Kostendeckung durch die Generierung von kurzfristigen Einnahmen<sup>80</sup>. Die in seinem Bericht zitierte Studie von *PIRA International* kommt jedoch zu dem Schluss, dass eine grenzenlose Freigabe der Daten für die Öffentlichkeit und die Privatwirtschaft größere monetäre Vorteile entstehen lassen würde, als aus Nutzungsgebühren erzielt werden kann. Da durch die Freigabe der Daten das Monopol der Behörden wegfallen würde, könnte sich eine verwertende Industrie um die Daten herum entwickeln. Diese Industrie würde zur Entstehung neuer Arbeitsplätze führen und dem Staat zu mehr Einnahmen aus direkten (Einkommenssteuer) und indirekten Steuern (Umsatzsteuer) verhelfen<sup>81</sup>. Das schließt natürlich auch eine auf diese Daten gestützte Medienberichterstattung mit ein.

Die Öffnung von öffentlichen Datenbeständen folgt außerdem dem substantiellen Argument der Open-Government-Bewegung (in Anlehnung an die Open-Data-Bewegung), dass die durch die Finanzierung aus Steuergeldern erhobenen Daten Allgemeingut sind und deswegen der Allgemeinheit kostenlos zur Verfügung gestellt werden sollten. Diese Forderung bezieht sich nur auf Daten des öffentlichen Sektors und schließt solche aus, die der Geheimhaltung unterliegen.

Mit den Portalen *Data.gov* und *data.gov.uk* sind die Regierungen der USA und von Großbritannien dieser Argumentation früh freiwillig gefolgt und Medien wie die Websites der *New York Times* und des britischen *Guardian* nutzen die freigegebenen Daten bereits intensiv für die eigene Berichterstattung (s. Kapitel 3.2).

In Deutschland, wo adäquate Portale noch fehlen, wird vor allem durch den Druck von Gesetzgebungen versucht, den Aufbau eigener Open-Government-Strukturen zu beschleunigen und voranzutreiben. Das aus der *PSI-Richtlinie* der EU hervorgegangene *Informationsweiterverwendungsgesetz*, das im Dezember 2006 in Kraft trat, soll dabei die „Weiterverwendung aller bei öffentlichen Stellen vorhandenen Informationen“<sup>82</sup> hierzulande regeln. Zwar sind bereits Open-Government-Projekte in Berlin und München aktiv<sup>83</sup>, die die klaren Vorteile offener Regierungsdaten aufzeigen und auch als Beleg für den Anstieg von Open-Government-Strukturen in Deutschland gesehen werden können, nach der Bundesregierung ist allerdings erst „bis 2013 der Aufbau einer zentral zugänglichen [...] Open-Data-Plattform vorgesehen“<sup>84</sup>.

79 *Data.gov* (o.J.): How were the datasets in Data.gov selected?, 18. Juli 2011

80 Vgl. Weiss (2002): *Borders in Cyberspace: Conflicting Public Sector Information Policies and their Economic Impacts*, S. 2

81 Vgl. ebenda, S. 5

82 Vgl. Bundesministerium der Justiz (2006): §1 Abs. 1 des Informationsweiterverwendungsgesetz

83 Vgl. z.B. *berlin.de* (o.J.): Open Cities; Matzat (2011), München will seinen Datenschatz heben, 19. Juni 2011

84 Bundesministerium für Wirtschaft und Technologie (2010): *Dresdner Vereinbarung*, S. 7

Trotz dieser für die Verhältnisse des Informationszeitalters langsamen Entwicklung, wird Open Government und die Open-Data-Bewegung in der Zukunft eine wichtige Rolle für die Gesellschaft spielen. Datenkompetenz vorausgesetzt, haben Journalisten anhand der vielschichtigen Veränderungen jetzt die Möglichkeit, die komplexer gewordene Gesellschaft besser erklären zu können, als jemals zuvor. Auf der einen Seite wird dadurch ein Zuwachs datengestützter Berichterstattung und von Datenvisualisierungen stattfinden. Auf der anderen Seite werden Journalisten zunehmend mit einem sich verändernden Berufsalltag konfrontiert werden. Denn da Daten stärker in den Fokus journalistischer Arbeit rücken, wird neben Recherche und Schreibarbeit auch die Arbeit mit Daten, einschließlich deren Aggregation, Management, Pflege und das Suchen nach Geschichten in den Daten eine essentielle Aufgabe von Journalisten werden. Diese Anforderungen an den Journalismus sammeln sich in dem neuen Arbeitsfeld des Datenjournalismus, der die journalistische Berichterstattung auf Grundlage von Daten als Fundament der Informations- und Wissensvermittlung fokussiert.

Parallel zu diesem neuen Arbeitsfeld werden Journalisten mehr und mehr gefordert sein, die verschiedenen Kommunikationskanäle wie Blogs und soziale Medien zu nutzen, deren Inhalte nach relevanten Informationen zu filtern und aus der Vielzahl von Informationen und Daten eine Geschichte zu kreieren. Dazu müssen sie die notwendigen Filtersysteme kennen und bedienen können.

Aufgrund dieser Veränderungen und angetrieben durch die Entwicklungen der letzten Jahrzehnte in Technik, Medien und Gesellschaft wird der Journalismus mit dem Aufkommen des Datenjournalismus große Veränderungen erleben und vor neue Herausforderungen gestellt. Diesen Herausforderungen müssen sich Journalisten sowie Medien stellen, wollen sie in ihrer Rolle als Vermittler von Nachrichten und Informationen an die Öffentlichkeit den Anforderungen des Informationszeitalters in den kommenden Jahren gerecht werden.

### 3. Daten im Journalismus

Die Arbeit mit Daten ist für Journalisten nichts Neues. Ob im Print- oder im Online-Journalismus, die Nutzung des Internets zur Recherche und die Einbeziehung von Studien, Fotos und Informationsgrafiken in die journalistische Berichterstattung existiert schon in Nachrichtenredaktionen seit der flächendeckenden Verbreitung von Computern und Internetzugang zu Beginn des neuen Jahrtausends<sup>85</sup>. Werden Daten in Form von Tabellen, Statistiken und Informationsgrafiken im klassischen Journalismus jedoch nur ergänzend zu den Beiträgen zur besseren Veranschaulichung des Inhalts genutzt<sup>86</sup>, verfolgt Datenjournalismus den Ansatz möglichst unverarbeitete Rohdaten als Grundlage jeglicher Berichterstattung zu nutzen. Das leitet sich auch aus der englischen Bezeichnung „*data driven journalism*“ ab, also ein von Daten angetriebener Journalismus.

Auf der einen Seite umschließt Datenjournalismus die Recherche in Datenbanken, das Sammeln, Analysieren und Auswerten sowie die leserfreundliche Visualisierung von Daten. Andererseits ist zur Gewährleistung einer größtmöglichen Transparenz auch die Herausgabe der Rohdaten, auf der ein Beitrag aufbaut, ein Grundsatz des Datenjournalismus. Damit folgt Datenjournalismus den Forderungen der Open-Data-Bewegung, die sich für mehr Transparenz und Freiheit im Umgang mit Daten einsetzt und die Verfügbarkeit von Daten in offenen, maschinenlesbaren Formaten fordert (s. Kapitel 2.3).

Ist das Ergebnis des klassischen Print- und Online-Journalismus die Veröffentlichung von Stories in Beiträgen der journalistischen Darstellungsformen, hat Datenjournalismus die Aufgabe Stories in Daten zu finden und diese in Datenbanken, Datenvisualisierungen und interaktiven Anwendungen zu publizieren. Diese im Journalismus neuen Darstellungsformen ermöglichen Datenjournalisten die Zusammenhänge in den Daten zu erkennen und die darin enthaltenen Stories ihren Lesern übersichtlich zu präsentieren. Gleichzeitig ermöglichen die Darstellungsformen aber auch den Lesern selbst mit den Daten aktiv zu arbeiten und Informationen in den Daten zu entdecken. Dabei kombiniert Datenjournalismus investigative Rechercheprozesse mit dem Wissen über Statistiken, Design und Programmierung<sup>87</sup> und eröffnet Datenjournalisten neue Möglichkeiten in der Berichterstattung.

Um diese zu erkennen, soll in diesem Kapitel gezeigt werden, welchen Einfluss Daten auf den Journalismus haben und welche Potentiale, aber auch Herausforderungen der Aufbau von Datenkompetenz bei Medien und Journalisten für einen funktionierenden Datenjournalismus mit sich bringt. Anhand von Beispielen nationaler und internationaler datenjournalistischer Berichterstattungen soll zudem der aktuelle Stand des Datenjournalismus in den Medien beleuchtet werden.

---

85 Vgl. Schneider/Raue (1998): Das neue Handbuch des Journalismus, S. 251 ff.

86 Vgl. Knieper (1995): Infographiken: Das visuelle Informationspotential der Tageszeitung, S. 5

87 Vgl. Bradshaw (2010): How to be a data journalist, 14. Juli 2011

## 3.1 Das journalistische Potential von Daten

In ihrer Rolle als Informanten und Kommentatoren des Zeitgeschehens sammeln Journalisten stetig Daten in Form von Informationen, um über aktuelle Ereignisse, Hintergründe, Entwicklungen oder Skandale berichten zu können. Sie recherchieren vor Ort, im Internet oder in Archiven und stellen Zusammenhänge zwischen den gesammelten Informationen her, die am Ende den Leser über die Medien erreichen und zu dessen öffentlicher Meinungsbildung beitragen. Der Einsatz von Daten kann dabei helfen, dass Journalisten diese Aufgaben schneller, umfassender und genauer ausführen können.

### 3.1.1 Datenrecycling

In seinem 2006 erschienenen Manifest zum Datenjournalismus weist der US-amerikanische Journalist und Webprogrammierer Adrian Holovaty darauf hin, dass eine Vielzahl der von Journalisten aggregierten Informationen bereits in einer strukturierten Form vorhanden sind<sup>88</sup>. Seine Feststellung erklärt er anhand eines Beispiels: So sind bei einer typischen Lokalberichterstattung über einen Brand gewisse Arten von Daten wie Tag, Zeitpunkt und Ort des Brandes, sowie Angaben über Opfer, Brandursache oder die Höhe des entstandenen Schadens immer vorhanden. Sortiert nach ihren Attributen und zusammengefasst in einem für Maschinen lesbaren Format würden die nach Holovaty so entstehenden Datensätze von Bränden es ermöglichen, dass die Daten vorangegangener und folgender Brände schnell miteinander verglichen werden können. Eingepflegt in eine Datenbank können die Daten außerdem auf einer interaktiven Karte mit verschiedenen Filteroptionen ausgegeben werden und offenlegen, in welchen Regionen es beispielsweise häufiger brennt<sup>89</sup>.

Die Aufbereitung in einer Datenbank ist bei kurzen Nachrichtentexten besonders einfach, da deren Informationsumfang gering ist und die Texte so eine schnelle Analyse ermöglichen. Nach Holovatys Beispiel einer Berichterstattung über einen Brand finden sich solche Texte oft im Lokaljournalismus. Aber auch im Sport- oder Wirtschaftsjournalismus, dessen Beiträge über die Entwicklung von Aktien und Unternehmen stark auf Zahlen und Fakten beruhen, lassen sich die Daten gut strukturieren und in Datenbanken aufbereiten<sup>90</sup>.

Anstatt also die vielen bereits gesammelten Daten lediglich in Nachrichtenarchiven einzulagern, sollten Medien ihre Archive anzapfen, die Datenattribute extrahieren, in Datenbanken integrieren und als Quelle für weitere Berichterstattungen nutzen. Vor dieser Wiederverwendung haben die meisten Daten der Tagespresse mit der Zeit stetig an Aktualität und damit an Wert verloren. In eine Datenbank eingepflegte, maschinenlesbare Daten, die sich dauerhaft mit aktuellen Berichterstattungen kombinieren lassen, sind dagegen eine verlustfreie Ressource für die Medien. Durch Datenrecycling ergeben sich daher neue und vor allem kostengünstige Recherchequellen für die Medien.

---

<sup>88</sup> Vgl. Holovaty (2006): A fundamental way newspaper sites need to change, 14. Juli 2011

<sup>89</sup> Vgl. ebenda

<sup>90</sup> Vgl. ebenda

### 3.1.2 Beschleunigung von Rechercheprozesse

In Archivmaterial nach Daten und Fakten zu recherchieren, ist für Journalisten ein mühsamer und zeitraubender Prozess. Zwar spricht der deutsche Journalist Walter von La Roche davon, dass „[i]n Zeitungshäusern zum Beispiel [...] alle Zeitungsartikel digitalisiert und gut recherchierbar [vorliegen]“<sup>91</sup>, aber im Vergleich der aktuellen Möglichkeiten von großen Zeitungsarchiven – im Hinblick auf deren Einsatz im Datenjournalismus – ergibt sich ein anderes Bild: Beispielsweise bietet die *Sueddeutsche Zeitung* in ihrem *MedienPort* genanntem Archiv mit Schlagworten, Dossiers und Vorschlägen zu ähnlichen Themen nützliche Features, um Artikel zu finden, aber ein Datenarchiv steht nicht zur Verfügung. So stehen am Ende einer Suche nur Artikel als Textform zur Verfügung. Darin enthaltene Informationen zu extrahieren kostet Journalisten viel Zeit bei der Recherche. Denn auch wenn das Material bereits digitalisiert als Text in einem redaktionellen Content-Management-System hinterlegt ist, muss es von Journalisten manuell durchsucht werden, da sich Text als Ganzes nicht nach gewünschten Kriterien filtern oder strukturieren lässt.

Die Arbeit mit Daten, die dagegen in einer Datenbank strukturiert worden sind, ermöglicht Journalisten daher eine Beschleunigung des Rechercheprozesses. Durch den Einsatz von Software wie dem kostenlosen *Google Spreadsheets*, Tabellenkalkulationsprogrammen wie *Microsoft Excel* oder mithilfe von Datenbankverwaltungssystemen wie dem freien *MySQL*, lassen sich die gesammelten Datenmengen schnell nach den gesuchten Kriterien filtern und relevante Informationen finden. Auf diese Weise steht Journalisten mehr Zeit für eine Analyse der Daten und eine Aufbereitung in Datenvisualisierungen zur Verfügung. Da die Informationen aus den eigenen Archiven stammen, sind die Daten als zuverlässige Quelle zudem bereits verifiziert.

### 3.1.3 Zusammenhänge aufdecken

Durch den Vorteil, dass „strukturierte[...] Informationen [...] mit anderen Datenbanken verknüpft und automatisiert abgerufen werden [können]“<sup>92</sup>, lassen sie sich einfach filtern und in übersichtlichen Grafiken ausgeben. So visualisiert, können Muster oder Anomalien in Form grafischer Berge und Täler besser erkannt und Zusammenhänge zwischen den Daten aufgedeckt werden, die vorher nicht deutlich waren.

Eine Datenvisualisierung der *New York Times* aus dem Jahr 2010 zeigt das Potential für den Journalismus: Bei der Analyse und dem Vergleich von US-Haushaltsprognosen seit 1980 kam zum Vorschein, dass 80 Prozent der Vier-Jahres-Prognosen optimistischer waren, als sich der Haushalt dann in der Realität entwickelt hatte<sup>93</sup> (s. Abb. 4). Durch den Vergleich der Prognose-Daten mit den Entwicklungsdaten und einer anschließenden Visualisierung konnte dieser Umstand aufgedeckt und der Öffentlichkeit anschaulich präsentiert werden. Anhand dieses neu gewonnenen Wissens aus alten Daten ist es den Medien zudem möglich, zukünftige Prognosen besser einschätzen und kritischer reflektieren zu können.

---

91 La Roche (1975): Einführung in den praktischen Journalismus, S. 15

92 Matzat (2011): Daten sind der Treibstoff des Journalismus, 14. Juli 2011

93 New York Times, (2010): Budget Forecasts, Compared With Reality, 14. Juli 2011

Published: February 2, 2010

## Budget Forecasts, Compared With Reality

Just two years ago, surpluses were predicted by 2012. How accurate have past White House budget forecasts been?

1 2 3 4 5 6 NEXT ▶

### Reasons for error

Budget forecasts require assumptions about both policy and the economy, and both can turn out to be spectacularly wrong.

In 2008, for example, the unemployment rate in 2012 was assumed to be 4.8 percent. The most recent assumption is 8.2 percent.



Abbildung 4:

Datenanalyse der US-Haushaltsprognosen seit 1980

Quelle:

New York Times, Budget Forecasts

### 3.1.4 Exklusivität

Aus den Archiven recycelte und aufbereitete Daten sind ein Schatz „unverbraachter“ Informationen. Da sich die alten Daten mit neuen verbinden lassen, entstehen unter Einsatz geringer Mittel neue Daten mit noch unveröffentlichten Zusammenhängen und Informationen. Durch eigene Datenvisualisierungen und journalistische Beiträge aufbereitet, können die neuen Inhalte von Medienhäusern exklusiv publiziert werden. Dadurch steigt auch einerseits die USP der Medien gegenüber ihrer Konkurrenz, andererseits werden sie unabhängiger von Datenzulieferern wie beispielsweise der *dpa-infocom* oder dem *dapd-Grafikdienst*. So ist Datenrecycling auch ökonomisch für Medien vorteilhaft.

### 3.1.5 Leserintegration durch Crowdsourcing

Für die Analyse großer Datenmengen, die Medien zum Beispiel von Enthüllungsplattformen wie *Wikileaks* erhalten oder über Open-Government-Kanäle der Öffentlichkeit freigegeben wurden (s. Kapitel 2.3.3), setzen Medien vermehrt auf Crowdsourcing (s. Kapitel 3.3.1 und 3.3.5). Die aktive Einbeziehung unbezahlter Arbeitskräfte führt einerseits dazu, dass die Daten schneller analysiert werden können und Journalisten früher Zugang zu den wichtigen Informationen in den Daten haben. Andererseits ermöglicht Crowdsourcing auch den Lesern selbst ein Teil der Medienberichterstattung zu werden. Durch die Nutzung solcher Synergien können Medien ihre Leser stärker an sich binden und ihre Arbeitsleistung erhöhen. Eine solche Integration der Leser in den Rechercheprozess ist jedoch an die Existenz von Werkzeugen gekoppelt, die Lesern ihre Arbeit erleichtern oder überhaupt erst ermöglichen.



Mit dem Aufruf „*Suchen Sie selbst und melden Sie sich bei uns!*“ versucht die *tageszeitung* (Taz) aus Berlin zum Beispiel seit Juni 2011 ihre Leser zum Crowdsourcing zu motivieren<sup>94</sup>. Mithilfe eines zur Verfügung gestellten Recherche-Suchwerkzeugs können die Leser alle Parteispenden über 10.000 Euro von 1994 bis 2009 durchsuchen<sup>95</sup>. Die verschiedenen Filteroptionen des Werkzeugs erleichtern ihnen dabei die Suche. Das Werkzeug wurde ergänzend zur Parteispenden-Karte 2009 von der *Taz* online bereitgestellt<sup>96</sup>. Die Daten stammen aus dem im Februar 2011 veröffentlichten Rechenschaftsbericht des Bundestags<sup>97</sup>.

### 3.1.6 Verdichtetes Storytelling

Es war immer die Aufgabe des Journalismus Informationen zu sammeln, zu kombinieren und dem Leser als aktuelle Beiträge zu präsentieren. Diese Aufgabe wird sich auch durch den Datenjournalismus nicht ändern. Was sich jedoch ändern wird, sind die Möglichkeiten, wie die Beiträge zukünftig präsentiert werden. Denn unter Einsatz digitaler Technologien können heute nicht nur intensivere, multimediale Berichterstattungen durch Fotostecken, Podcasts, Informationsgrafiken und Videos angeboten werden, mithilfe von interaktiven Datenvisualisierungen werden die Beiträge auch erlebbarer präsentiert.

Alan McLean, Entwickler im *Interactive News Technology*-Team der *New York Times* (NYT), zeigt dieses verdichtete Storytelling am Beispiel der Berichterstattung der NYT über die afghanischen *War Diaries*, die durch *Wikileaks* enthüllt wurden<sup>98</sup>: So besteht das Kernstück der Berichterstattung aus drei Artikeln, die den Krieg durch Text-Auszüge der Kriegsdokumente thematisch verschieden beschreiben und nutzerfreundlich zu den Originaldokumenten verlinken. Darüber hinaus verfasste die NYT zwei Artikel über den Ursprung der Daten und den Rechercheprozess und erstellte einen Bereich, in dem ausgewählte und mit Zusatzinformationen ausgestattete Dokumente durchgeschaut werden können. Die Artikel selbst werden noch mit aufbereiteten Karten, einer geschichtlichen Entwicklung des Landes, Fotos und Audio-Beiträgen begleitet<sup>99</sup>. Alle Informationen sind dabei übersichtlich auf einer Themenseite mit eigener Navigation zusammengefasst<sup>100</sup>. Das durch die Vielzahl der eingesetzten Medien so entstehende Bild des Krieges ermöglicht durch die Verflechtung von Datenjournalismus und Online-Journalismus eine intensivere Berichterstattung.

94 taz.de (2011): Parteispenden Recherche 1994-2009 - ab 10.000 Euro, 14. Juli 2011

95 Vgl. ebenda

96 Vgl. taz.de (2011): Parteispenden Watch 2009 - ab 10.000 Euro, 14. Juli 2011

97 Vgl. ebenda, siehe „Info“, 14. Juli 2011

98 Vgl. New York Times (2010): Piecing Together the Reports, and Deciding What to Publish, 14. Juli 2011

99 Vgl. McLean (2010): Data Driven Journalism - Telling Stories Online, S. 30, 14. Juli 2011

100 Vgl. Mazetti et al. (2010): Pakistan Aids Insurgency in Afghanistan, Reports Assert, 14. Juli 2011

## 3.2 Daten im Einsatz

### 3.2.1 Spesenskandal des Guardian



**guardian.co.uk** Your search terms... Entire site Search

News Sport Comment Culture Business Money Life & style Travel Environment TV Blogs Data Mobile Offers Jobs

Investigate your MP's expenses: Home → Labour

### Gordon Brown MP's expenses

[Start reviewing pages](#)

**Documents**

Document	Total reviewed	Progress
IEP 2006/07	102 of 102 pages reviewed	<div></div>
Additional Costs Allowance 2004/05	32 of 32 pages reviewed	<div></div>
Incidental Expenses Provision 2004/05	91 of 91 pages reviewed	<div></div>
Additional Costs Allowance 2005/06	36 of 36 pages reviewed	<div></div>
Incidental Expenses Provision 2005/06	120 of 120 pages reviewed	<div></div>
Additional Costs Allowance 2006/07	48 of 48 pages reviewed	<div></div>
Additional Costs Allowance 2007/08	62 of 62 pages reviewed	<div></div>
Communication Allowance 2007/08	3 of 3 pages reviewed	<div></div>
Incidental Expenses Provision 2007/08	66 of 66 pages reviewed	<div></div>

**Recent investigations**

- anon-1317
- zebedee
- anon-2161
- anon-1671
- danvers

Labour MP for Kirkcaldy & Cowdenbeath  
[Guardian politics profile](#)

**Abbildung 5:**

Spesenabrechnungen des damaligen  
Premierministers Gordon Brown

**Quelle:**

guardian.co.uk, MP's expenses

Die Berichterstattung des britischen *Guardian* im Frühjahr 2009 über den Spesenskandal des *Houses of Parliaments* ist ein Pionierstück datenjournalistischer Arbeit. Denn erstmals im Investigativ-Journalismus wurde eine umfangreiche Berichterstattung auf Basis eines Datensatzes durch die aktive Mitarbeit der Öffentlichkeit über das Internet ermöglicht. Der Skandal entstand im Mai 2009 mit der Exklusivveröffentlichung von Einzelheiten aus den Spesenabrechnungen von Parlamentsmitgliedern durch den *Daily Telegraph*, dem die Dokumente von Regierungsmitarbeitern zugespielt wurden<sup>101</sup>. Einen Monat später gab das Parlament die Dokumente auch für die Öffentlichkeit frei<sup>102</sup>. Da dem *Guardian* jedoch die personellen Ressourcen fehlten den über 450.000 Seiten starken Datensatz<sup>103</sup> eigenständig zu analysieren, veröffentlichte er die Dokumente auf seiner Website und bat seine Leser bei der Suche nach relevanten Informationen um aktive Mithilfe<sup>104</sup>. Zusätzlich stellte der *Guardian* eine Webanwendung zur Verwaltung und Analyse bereit (s. Abb. 5). So konnten unter dem Einsatz von über 28 000 Lesern und Helfern fast 50 Prozent der Dokumente ausgewertet werden<sup>105</sup>. Im Anschluss publizierte der *Guardian* die auf Grundlage der Daten

101 Vgl. o.V. (2009): Expenses: How MP's expenses became a hot topic, 14. Juli 2011

102 Vgl. Rogers (2009): How to crowdsource MPs' expenses, 14. Juli 2011

103 Vgl. guardian.co.uk (2009): Investigate your MP's expenses, 14. Juli 2011

104 Vgl. Rogers (2009): How to crowdsource MPs' expenses, 14. Juli 2011

105 Vgl. guardian.co.uk (2009): Investigate your MP's expenses, 14. Juli 2011



gewonnenen Informationen in klassischen Darstellungsformen des Journalismus und legte damit einen der größten Politskandale Großbritanniens offen<sup>106</sup>.

Die Aufbereitung des Spesenskandals des *Guardian* zeigt das Potential von Datenjournalismus für die investigative Berichterstattung. Viele führende Online-Medien wie die *New York Times* oder das *Wall Street Journal* (s. Kapitel 5) haben dieses Potential erkannt und nutzen seit 2010 vermehrt Daten als Grundlage für ihre publizistische Arbeit. Aber auch in deutschen Medien steigt die Aufmerksamkeit auf das Thema durch einzelne Projekte.

### 3.2.2 Verräterisches Handy von ZEIT ONLINE

So veranschaulicht das Projekt *Verräterisches Handy* von *ZEIT ONLINE* (s. Abb. 6) die negativen Folgen des 2007 verabschiedeten Gesetzes zur Vorratsdatenspeicherung<sup>107</sup>, durch das „*Telefon- und Internetverbindungsdaten [...] ein halbes Jahr lang gespeichert*“<sup>108</sup> werden dürfen. Um die Auswirkungen des Gesetzes auf die Privatsphäre der Bürger und den Informationsgehalt der gespeicherten Daten darzustellen, wurden die gespeicherten Handy-Daten von einem halben Jahr des Grünen-Politikers Malte Spitz in einer interaktiven Datenvisualisierung verarbeitet. Angereichert mit weiteren öffentlichen Informationen von Spitz' Twitter-Account und Blog, sowie Websites, in denen der Politiker erwähnt wird, entstand so ein komplexes Profil seines Lebens, das seine Gewohnheiten, Aufenthaltsorte und Reiserouten offen legt<sup>109</sup>.

Das Projekt wurde mithilfe eines *Mashups*<sup>110</sup> auf Basis von *Google Maps* realisiert. Um für eine bessere visuelle Trennung zwischen der Karten-Ebene und der Ebene der visualisierten Daten zu sorgen, wurde das Kartenmaterial von *Google Maps* noch farblich angepasst (Weiß für Landstriche, Hellgrau für Grünflächen) und inhaltlich reduziert (Ausblendung von Straßennamen und weniger wichtigen Ortsnamen und -teilen). Außerdem wurde die Visualisierung durch verschiedene Möglichkeiten der Interaktion und Navigation nutzerfreundlich erweitert. Die der Visualisierung zugrunde liegenden Rohdaten wurden in einer Tabellendatei mit offenem Dateiformat für die Nutzer zu Zwecken der eigenen Analyse und Recherche frei zugänglich gemacht.

Ergänzt wurde die Berichterstattung mit dem Artikel von Kai Biermann „*Was Vorratsdaten über uns verraten*“<sup>111</sup>. In dem Artikel wurden der thematische Hintergrund und die Erkenntnisse aus der Datenvisualisierung erläutert und auf weitere Möglichkeiten der Überwachung durch die Verwendung gespeicherter SMS- und Telefondaten eingegangen, da diese Daten zum Schutz der Privatsphäre von Malte Spitz nicht mit in das Projekt eingebunden wurden<sup>112</sup>.

Durch die eingängige Visualisierung und die Kombination von klassischer und datengestützter Berichterstattung verdeutlichte *ZEIT ONLINE* die Brisanz der Vorratsdatenspeicherung wesentlich stärker, als das durch eine Berichterstattung allein auf Textbasis

106 Vgl. z.B. Webster (2009): Parliament's darkest day: MPs suspended and Michael Martin at risk; Coates (2009): Behind the story: MPs turn fire on Andrew Walker, man who approved claims, 14. Juli 2011

107 ZEIT ONLINE (2011): Verräterisches Handy, 14. Juli 2011

108 o.V. (2007): Köhler unterzeichnet Gesetz zur Vorratsdatenspeicherung, 14. Juli 2011

109 Vgl. Biermann (2011): Was Vorratsdaten über uns verraten, S. 1, 14. Juli 2011

110 Ein Mashup ist die Kombination bestehender Inhalte zur Erstellung neuer Inhalte.

111 Vgl. Biermann (2011): Was Vorratsdaten über uns verraten, S. 1, 14. Juli 2011

112 ebenda, S. 2, 14. Juli 2011

möglich gewesen wäre. Für diese erfolgreiche Arbeit im Datenjournalismus erhielt das Projekt den *Grimme Online Award 2011*, da es nach Ansicht der Jury einen „wichtigen Beitrag zur Kultivierung dieses Genres geleistet“ hat, „das so nur im Web möglich ist“<sup>113</sup>.

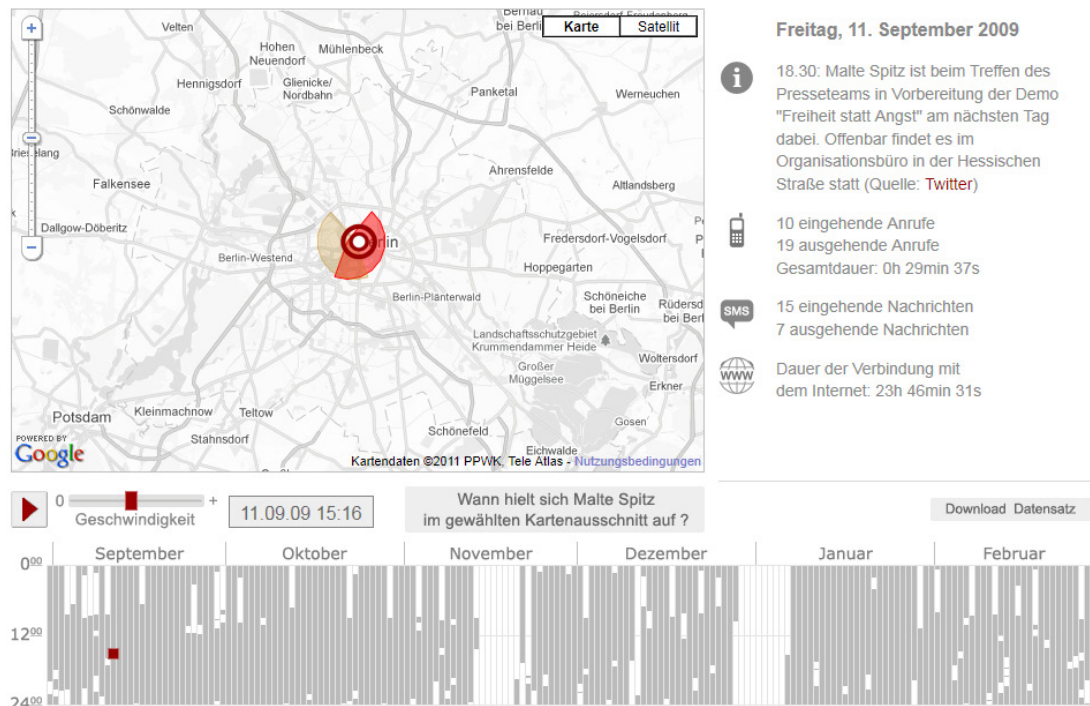


Abbildung 6:

Datenvisualisierung der Auswirkungen  
der Vorratsdatenspeicherung

Quelle:

ZEIT ONLINE, Verräterisches Handy

### 3.2.3 Fluglärm-Karte der Taz

Ein weiteres Beispiel für den Einsatz von erklärenden Datenvisualisierungen in deutschen Medien ist die Fluglärm-Karte der *Taz*<sup>114</sup> über die zu erwartende Lärmentwicklung vor Ort und im Umland des sich im Bau befindlichen Flughafens Berlin Brandenburg. Als Grundlage des Projekts wurde wie bei *Verräterisches Handy* ein Mashup einer angepassten Version von *Google Maps* genutzt. Die Daten stammen aus der Veröffentlichung „Abschätzung der Fluglärmbelastungen durch BBI-Flugrouten“ des *Ministerium für Infrastruktur und Landwirtschaft Brandenburgs* und werden zudem in maschinenlesbarer Form zum Download angeboten<sup>115</sup>.

Die Visualisierung ist in die aktuelle Situation am Flughafen, den zwei eingebrachten Vorschlägen zu den Flugrouten und einer ergänzenden Informationsgrafik über Flughöhen unterteilt (s. Abb. 7). Kernstück der Datenvisualisierung sind zwei Visualisierungen der von den jeweiligen Vorschlägen ausgehenden Dauerschallpegel und der maximal erreichbare Schallpegel beim Landen und Starten von Düsenjets. Der Dauerschallpegel wurde durch Isolinien dargestellt, der maximal erreichbare Schallpegel durch Punkte auf den Isoflä-

<sup>113</sup> Grimme Online Award (2011): Begründung der Jury [Verräterisches Handy], 14. Juli 2011

<sup>114</sup> taz.de (2011): Fluglärm-Karte BBI-Airport Berlin, 14. Juli 2011

<sup>115</sup> Vgl. ebenda, siehe „Hintergrundinfo“, 14. Juli 2011

chen<sup>116</sup>. Da beide Datensätze die gleiche Farbskala benutzen und die Isoflächen durch den Einsatz von Transparenzen die darunter liegende Umgebung nicht komplett verdeckt, bleibt die Datenvisualisierung übersichtlich und die darauf vermittelten Informationen können vom Nutzer durch den starken Qualitätskontrast schnell aufgenommen werden. Angereichert wird die datengestützte Berichterstattung durch thematisch passende Artikel, die auf der Karte als Icon an den Orten erscheinen, auf die sich die Artikel inhaltlich beziehen.

Durch die Angabe des Dauerschallpegels beim Klick auf die Isoflächen, ist die Visualisierung nicht nur für die Allgemeinheit interessant, sondern vor allem auch für die Anwohner im Umfeld. Denn Immobilienbesitzer und Hausbauer haben durch die Karte Zugriff auf sehr genaue Ortsdaten, die ihnen Aufschluss darüber geben, inwieweit ihr Grundstück vom Fluglärm betroffen sein wird<sup>117</sup>. Hier zeigt sich das Potential solcher Projekte für einen neuen Lokaljournalismus, dessen Berichterstattung auf hyperlokalen Daten aufbaut.

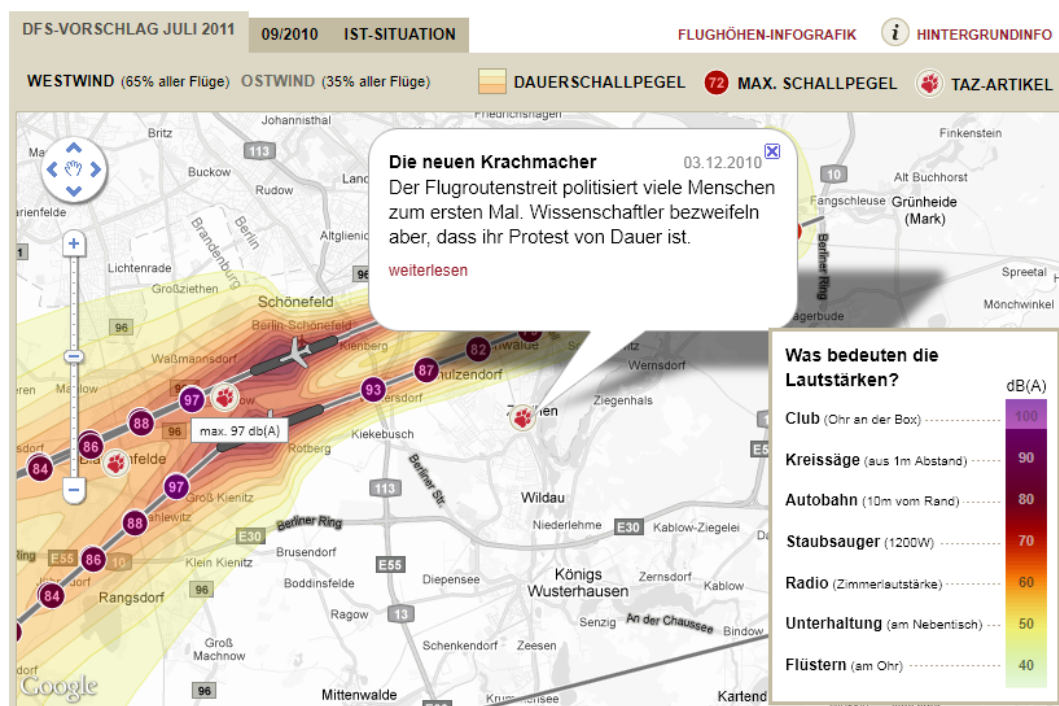


Abbildung 7:

Fluglärmkarte für den geplanten Flughafen BBI

Quelle:

die tageszeitung, Fluglärm BBI

### 3.2.4 Status quo im Datenjournalismus

Obwohl die Projekte *Verräterisches Handy* und die Fluglärm-Karte gute Beispiele für datenjournalistische Berichterstattungen in Deutschland sind, legen sie sogleich ein vorhandenes Defizit der Medien an Datenkompetenz und geeignetem Personal offen. Denn keines der Projekte wurde von den Medien selbst entwickelt. In beiden Fällen war für die Idee und die Umsetzung das Berliner Datengestaltungs-Büro *OpenDataCity* zuständig, das die Projekte erst eigenständig entwickelte und später interessierten Medien anbot<sup>118</sup>.

<sup>116</sup> Vgl. taz.de (2011): Fluglärm-Karte BBI-Airport Berlin, 14. Juli 2011

<sup>117</sup> Vgl. Matzat (2011): Fluglärmkarte BBI: Making of, 14. Juli 2011

<sup>118</sup> Vgl. OpenDataCity (2011): Fluglärmkarte BBI;  
OpenDataCity (2011): Vorratsdatenspeicherung; 14. Juli 2011

Um in Zukunft eigene Projekte aufbauen zu können, deren Möglichkeit sich zum Beispiel aus recherchiertem oder exklusivem Datenmaterial ergibt, benötigen Medienhäuser eigene Datenjournalisten in ihren Redaktionen. Auch hier gilt das Engagement des *Guardian* als Musterbeispiel. So hat die Zeitung den Daten mit dem *Datastore* ein ganzes Ressort auf seiner Website gewidmet, das sogar zu externen Blogartikeln verlinkt<sup>119</sup>. Zusätzlich eröffnete der *Guardian* mit dem *Datablog* eine Plattform, auf der datengestützte Themen auch außerhalb der täglichen Nachrichtenagenda veröffentlicht werden<sup>120</sup>. Unter den Autoren des *Datablog* finden sich sowohl Redakteure des *Guardian*, als auch freie Datenjournalisten und -spezialisten, die exklusiv verschiedene Themen und Visualisierungen mit Datenhintergrund publizieren. Für dieses einmalige Engagement erhielt der *Guardian* für den *Datablog* den Newspaper Award 2011<sup>121</sup>.

Von so einer Auszeichnung sind deutsche Medien noch weit entfernt – vor allem, da Datenjournalisten in Deutschland noch knapp sind. Die freie Medien- und Marketingjournalistin Ulrike Langner, die sich stark im Bereich Datenjournalismus engagiert, geht davon aus, dass „*sich bisher nur rund ein Dutzend Journalisten und Programmierer intensiv mit dem Thema*“<sup>122</sup> beschäftigen. So ist die Datenjournalismus-Szene in Deutschland noch sehr überschaubar, was den Medienhäusern den Aufbau von Datenkompetenz nicht erleichtert und Veröffentlichungen eigener, unabhängiger Daten-Berichterstattungen noch verhindert.

### 3.2.5 Die Plagiatsaffäre, ein Selbstläufer

Trotz der fehlenden Datenexperten in Deutschland sind viele Internetnutzer in der Arbeit mit Daten sehr aktiv. Einerseits leisten sie durch Crowdsourcing häufiger investigative Datenarbeit, andererseits beeinflussen immer öfter Einzelpersonen mit ihren Enthüllungen das Agenda-Setting der Leitmedien (s. Kapitel 2.2). Demzufolge untergraben Internetnutzer immer mehr die Deutungshoheit und den Einflussbereich der Medien. Die Kommunikationswissenschaftlerin Miriam Meckel nimmt an, dass sich dieser Prozess noch verstärken wird und rät Journalisten deshalb „*das Netz als Recherche-, Kollaborations- und Kommunikationsplattform*“<sup>123</sup> zu nutzen und zu akzeptieren, dass Journalismus heute „*in Prozesse der kollaborativen Informations- und Geschichtenproduktion*“<sup>124</sup> eingebettet ist und „*Hinweise auf investigative Stories [...] sich nicht mehr nur aus journalistischer Nachforschung, sondern durch Tipps aus den sozialen Netzwerken*“<sup>125</sup> ergeben. Nach Meckel entstehen Geschichten also nicht länger durch die Arbeit einzelner Journalisten, sondern durch das Zusammenwirken einer Vielzahl unterschiedlicher Medienakteure und Quellen.

Beispielhaft für die Rolle der Medien in diesem Geflecht ist die Plagiatsaffäre um den ehemaligen Bundesverteidigungsminister Karl-Theodor zu Guttenberg: Am 16. Februar 2011 berichtete die Süddeutsche Zeitung erstmals über Plagiate in der Doktorarbeit von Karl-

119 Vgl. guardian.co.uk (o.J.): Data blogosphere: today's picks, 14. Juli 2011

120 Vgl. guardian.co.uk (o.J.): Datablog, 14. Juli 2011

121 Vgl. Gunter (2011): Newspaper Awards 2011: Guardian takes home digital prizes for website and datablog, 14. Juli 2011

122 Langner (2011): Daten rücken in den Fokus, Fachjournalist 2/2011, S. 5

123 Meckel (2011): Journalisten an der Crowdsourcing-Front, S. 2, 14. Juli 2011

124 ebenda

125 ebenda, S. 1



Theodor zu Guttenberg<sup>126</sup>, nachdem sich deren Entdecker, der Jura-Professor Andreas Fischer-Locarno, an die Zeitung gewandt hatte. Um seine Feststellungen zu belegen, stellte er zudem eine Vorabversion seiner Rezension über die Doktorarbeit für die Zeitschrift „Kritische Justiz“ den Medien zur Verfügung<sup>127</sup>. Noch am selben Tag entdeckten mehrere Medien weitere Plagiate<sup>128</sup>. Daraufhin wurde am 17. Februar unter Mithilfe des österreichischen Experten für Wissenschaftsplagiate Stefan Weber das Projekt *GuttenPlag Wiki* gestartet<sup>129</sup>, das für seine Arbeit den *Grimme Online Award 2011* erhielt. Durch die Mitarbeit von über eintausend Freiwilligen<sup>130</sup> deckte das Projekt auf über 90 Prozent der Seiten der Doktorarbeit plagiierte Textpassagen auf<sup>131</sup> und trug einen maßgeblichen Anteil an der weiteren Aufklärung der Affäre<sup>132</sup>. Diese gipfelte am 1. März 2011 im Rücktritt des Bundesverteidigungsministers aufgrund des zunehmenden Drucks<sup>133</sup>. So hat die Plagiats-Affäre bewiesen, dass die für eine journalistische Berichterstattung relevante Recherchearbeit auch im Netz stattfinden und als Ergebnis des Zusammenwirkens von Privatpersonen, Medien und der Masse in kürzester Zeit viel bewegen kann.

Für die deutschen Leitmedien war die Affäre vor allem ein Beweis für die Grenzen ihrer Leistungsfähigkeit. Konnte der *SPIEGEL* in seiner fünf Tage nach Bekanntwerden der Plagiate erschienen Ausgabe gerade einmal 60 Plagiate aufdecken, hatten die Mitarbeiter des *GuttenPlag Wiki* bereits über 300 Plagiate entdeckt<sup>134</sup>. Auch hätte die Affäre ohne die Arbeit und genaue Dokumentation der vielen Freiwilligen einen anderen Ausgang genommen und wäre womöglich „versandet“, bewertet Christian Stöcker, Ressortleiter Netzwelt von *SPIEGEL ONLINE*, die Entwicklungen<sup>135</sup>. Es ist daher davon auszugehen, dass kollaborative Projekte in naher Zukunft eine wichtige Rolle bei der Arbeit mit umfangreichen Texten und Datensätzen spielen werden. Die Jury des *Grimme Online Award* ist zumindest davon überzeugt, dass das Projekt *GuttenPlag Wiki* verdeutlicht „welche Möglichkeiten das Web generell für gemeinsames Arbeiten bietet“<sup>136</sup>. Und auch der Medienökonom Robin Meyer-Lucht sieht deutlich das Potential kollaborativer Zusammenarbeit:

„Das Mitmach-Recherche-Netz hat hier die Recherche-Leitmassenmedien überflügelt“

(Robin Meyer-Lucht, deutscher Medienökonom)<sup>137</sup>

Um in der Lage zu sein, solche Investigativ-Projekte bei fehlenden Impulsen aus der Bevölkerung auch selbst anstoßen zu können, müssen Journalisten lernen mit der Masse zu arbeiten, diese anzuleiten, Werkzeuge bereitzustellen und die Ergebnisse zu koordinieren und zu bewerten.

126 Vgl. Preuß/Schultz (2011): Guttenberg soll bei Doktorarbeit abgeschrieben haben, 14. Juli 2011

127 Vgl. Schnabel (2011): „Ich wollte es nicht glauben“, 14. Juli 2011

128 Vgl. z.B. o.V (2011): Bei diesen Autoren könnte Guttenberg geklaut haben; Georgi (2011): Anfang bei F.A.Z. abgeschrieben, 14. Juli 2011

129 Vgl. Neuroth (2011): „Schwarmintelligenz“ im Kampf gegen Plagiate, 14. Juli 2011

130 Vgl. Grimme Online Award (2011): Preisträger 2011, 14. Juli 2011

131 Vgl. GuttenPlag (o.J.): 14. Juli 2011

132 Vgl. Fischer/Medick (2011): Copy, paste, delete, 14. Juli 2011

133 Vgl. ebenda

134 Vgl. Pohlmann (2011): Schwarm und Schwärmer, 14. Juli 2011

135 Stöcker (2011): Netz besiegt Minister, 14. Juli 2011

136 Vgl. Grimme Online Award (2011): Begründung der Jury [GuttenPlag Wiki], 14. Juli 2011

137 Vgl. Pohlmann (2011): Schwarm und Schwärmer, 14. Juli 2011

## 3.3 Herausforderungen für den Journalismus

### 3.3.1 Mut zum Datenjournalismus

Vielen Zeitungen im Online-Bereich fehlt noch der Mut sich Kompetenz im wachsenden Feld des Datenjournalismus anzueignen. Mut, der nach Entwicklungsredakteur Sascha Venohr von *ZEIT ONLINE* die Medien und Redaktionen aktuell noch daran hindert *„Geschichten nicht wie bisher auf sechs Seiten Text, sondern mit den neuen Darstellungsformen zu erzählen“*<sup>138</sup>. Da in Deutschland geeignete Ausbildungsangebote erst langsam entstehen – die Henri-Nannen-Schule baut zum Beispiel erst seit 2011 den Lehrbereich Datenjournalismus auf<sup>139</sup> – und sich erst wenige deutsche Journalisten mit Datenjournalismus auseinandersetzen, müssen die Medien selbst beginnen Datenkompetenz in ihren Redaktionen aufzubauen, mit datenjournalistischen Berichterstattungen zu experimentieren und auch Nachwuchsjournalisten aktiv zu fördern. Vor allem, da in Deutschland geeignete Förderprogramme wie die US-amerikanische *Knight Foundation* zur Finanzierung von Innovationen im Journalismus fehlen<sup>140</sup>.

Aber auch Journalisten brauchen mehr Mut. Komplexe Datenvisualisierungen mögen einschüchternd wirken und einen erheblichen Anteil an Programmierkenntnissen voraussetzen. In Wirklichkeit sind sich nach Ulrike Langner *„alle [Experten] [...] einig, dass spezielle Programmierkenntnisse weitaus weniger wichtig sind als der richtige Umgang mit dem Tabellenkalkulationsprogramm Excel“*<sup>141</sup>. Von Vorteil sind außerdem *„Grundkenntnisse statistischer Zusammenhänge und dass man Datenmaterial auf seine Verwertbarkeit hin zuverlässig beurteilen kann“*<sup>142</sup>. Da diese Voraussetzungen schon heute von vielen Online- und Printjournalisten erfüllt werden, die es gewohnt sind, fremde Statistiken auf ihren Inhalt und ihren Wahrheitsgehalt zu überprüfen, ist der Einstieg in den Datenjournalismus weniger hürdenreich, als er auf den ersten Blick erscheint.

### 3.3.2 Interdisziplinäre Datenteams aufbauen

Trotz dem einfachen Einstieg sollten Journalisten trotzdem die Fähigkeit beherrschen in einem interdisziplinären Team mit Programmierern, Statistikern und Designern arbeiten zu können (s. Abb. 8). Denn *„[e]in Journalist alleine kann große Datensätze nicht sinnvoll bewältigen. Es braucht immer Leute, die aus verschiedenen Professionen mit verschiedenen Kompetenzen herangehen, um ein wirklich verwertbares Produkt herzustellen“*<sup>143</sup>, meint der Datenjournalist und Mitgründer von *OpenDataCity* Lorenz Matzat. Deswegen sollten Medien Datenteams aufbauen, die gemeinsam nach interessanten Daten suchen, sie auswerten, verarbeiten und in Beiträge verpackt präsentieren.

---

138 Langner (2011): Schaubild statt Klickstrecke, Medium Magazin 1-2/2011, S. 42

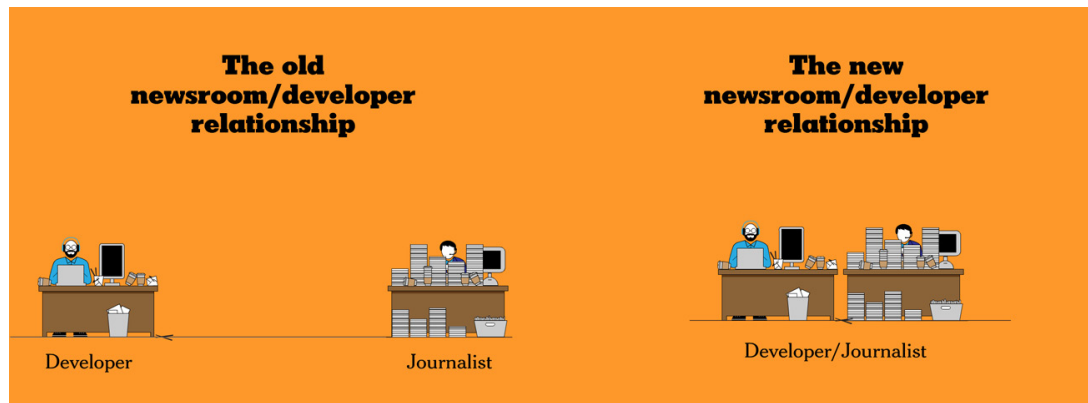
139 Vgl. Langner (2011): Daten rücken in den Fokus, Fachjournalist 2/2011, S. 5

140 Vgl. Beckedahl (2011): Datenjournalismus und die Zukunft der Berichterstattung, 14. Juli 2011

141 Langner (2011): Schaubild statt Klickstrecke, Medium Magazin 1-2/2011, S. 42

142 Langner (2011): Daten rücken in den Fokus, Fachjournalist 2/2011, S. 7

143 Langner (2011): „Man macht sich überprüfbar“, Medium Magazin 1-2/2011, S. 43

**Abbildung 8:**

Journalisten müssen lernen mit  
Spezialisten zusammenzuarbeiten

**Quelle:**

Allan McLean, Telling Stories With Data

Neben dem internen Aufbau von Datenteams besteht auch die Möglichkeit, auf bereits vorhandene Datenkompetenz externer Dienstleister für datenjournalistische Projekte zurückzugreifen. Eine solche Auslagerung bietet ökonomisch den Vorteil, dass keine Kosten für noch knappes Fachpersonal und Ressourcen anfallen und das eigene Risiko bei experimentellen Projekten minimiert wird. Nachteilig sind jedoch die verminderte Flexibilität durch die fehlende Integration in den redaktionellen Workflow und die fehlende Investition in die eigene Kompetenz und Unabhängigkeit.

### 3.3.3 Transparenter werden

Ein Grundsatz von Open Data ist die Freigabe von Rohdaten ohne Beschränkung. So können sie recycled und in Verbindung mit anderen Daten gebracht werden. Wollen Medien die Masse zur Analyse ihrer Daten über Crowdsourcing gewinnen, um aufbauend auf den so gewonnenen Erkenntnissen datengestützte Beiträge veröffentlichen zu können, müssen sie ihre Daten ohne Restriktionen frei zugänglich machen. Diese Quellen-Offenheit und Transparenz ist vielen Zeitungshäusern noch unangenehm, meint Lorenz Matzat, aber „[d]ie Leute müssen sich daran gewöhnen, sich in die Karten schauen zu lassen“<sup>144</sup>.

Die Transparenz nimmt nach Jeff Jarvis den Stellenwert einer neuen Objektivität ein<sup>145</sup>, die im klassischen Journalismus durch die Sorgfaltspflicht beschrieben wurde. So besagt Ziffer 2 des Pressekodex, dass im Journalismus „[z]ur Veröffentlichung bestimmte Informationen in Wort, Bild und Grafik [...] mit der nach den Umständen gebotenen Sorgfalt auf ihren Wahrheitsgehalt zu prüfen und wahrheitsgetreu wiederzugeben [sind]“<sup>146</sup>. Mit der Offenlegung der Quellen können Medien ihrer Sorgfaltspflicht nachkommen und Leser haben die Möglichkeit, die Fakten einer Berichterstattung selbst zu prüfen und nachzuvollziehen. Dadurch gewinnen die Medien einerseits an Transparenz, andererseits gewinnen sie auch

<sup>144</sup> Langner (2011): „Man macht sich überprüfbar“, Medium Magazin 1-2/2011, S. 43

<sup>145</sup> Vgl. Empson (2011): When It Comes To New Journalism, 'Transparency Is The New Objectivity', 14. Juli 2011

<sup>146</sup> Presserat (o.J.): Ziffer 2 - Sorgfalt, 14. Juli 2011

an Vertrauen in der Öffentlichkeit. Vor allem für Zeitungs-Websites ist das wichtig, da sie stark auf die Mitarbeit der Öffentlichkeit bei Crowdsourcing-Projekten angewiesen sind.

### 3.3.4 Vertrauen generieren

Seriöse Online-Medien werden sich verstärkt am Markt durchsetzen. Vertrauen spielt für Mediennutzer bei der Auswahl von Medien zunehmend eine Rolle. Davon geht zumindest eine Umfrage des *MedienExpertenPanel* im Auftrag der *Bauer Media Akademie* aus dem Jahr 2007 aus. Die teilnehmenden Medienexperten gaben an, dass die Bedeutung von Seriosität und Glaubwürdigkeit mit 56 Prozent noch vor der Möglichkeit zur Interaktion und Mitgestaltung von Inhalten steht (42 Prozent)<sup>147</sup>. Die zeitliche und örtliche Verfügbarkeit der Medien sahen die Experten mit 80 Prozent als bedeutendster Nutzungsfaktor für die Zukunft<sup>148</sup>. Das trifft vor allem auf Online-Medien zu. Durch die Offenlegung von Quellen und einer transparenten Berichterstattung in datenjournalistischen Projekten können Medien dieses Vertrauen aufbauen.

### 3.3.5 Journalisten als Kuratoren von Informationen

*„Curators help navigate readers through the vast ocean of content“*

(Josh Sternberg, US-amerikanischer Kommunikationsstrategie)<sup>149</sup>

Die wohl größte Herausforderung im Datenzeitalter, der sich Journalisten annehmen müssen, ist ein verstärktes Auftreten als Kurator von Medieninhalten. Zeitungen, Radio- und Fernsehsender, Pressestellen, Websites, Nachrichtenagenturen, Blogs und soziale Netzwerke – die verschiedenen Medienkanäle produzieren dauerhaft ein für Einzelpersonen undurchdringbares Aufkommen an Informationen. Zusätzlich veröffentlichen Wissenschaftler, Unternehmen und Regierungen mehr Daten. Da Journalisten ihre sinnstiftende Funktion und Deutungshoheit von Inhalten durch das Internet verlieren<sup>150</sup>, müssen sie sich ihrer Rolle anpassen und vermehrt andere Aufgaben übernehmen. Im Datenjournalismus werden Journalisten so zu Kuratoren relevanter Inhalte. So übernimmt beispielsweise der *Guardian* mit seinem *Datablog* die Aufgabe für seine Leser bei der Suche nach Informationen zu helfen, indem die Redaktion auf aktuelle und relevante Daten hinweist<sup>151</sup>. Gleichzeitig stellt der *Guardian* mit dem *DataBlog* auch Werkzeuge bereit, um seinen Lesern zu helfen die Daten zu analysieren<sup>152</sup>. Die Arbeit von Journalisten als aktive Informationsproduzenten rückt damit in den Hintergrund, ihre Bedeutung als Informationsintermediär und -manager, die Inhalte zwischen ihren Lesern und ihren Quellen vermitteln, wird wichtiger. Journalisten werden zu *„Aggregatoren oder Broker[n], die für die Sammlung relevanter Informationen im Netz oder als Schnittstelle zur Verbindung unterschiedlicher Communities zuständig sind“*<sup>153</sup> meint Miriam Meckel. Damit ist ein weitreichender Paradigmenwechsel im Journalismus verbunden.

147 Bauer Media Akademie (2007): Zukunftswerkstatt Medien: MedienExpertenPanel, S. 18

148 ebenda

149 Sternberg (2011): Why Curation Is Important to the Future of Journalism, 14. Juli 2011

150 Vgl. Meckel (2011): Journalisten an der Crowdsourcing-Front, S. 2, 14. Juli 2011

151 Vgl. Stray (2010): How The Guardian is pioneering data journalism with free tools, 14. Juli 2011

152 Vgl. Littledale/Baily (2010): Guardian News Traces, 14. Juli 2011

153 Meckel (2011): Journalisten an der Crowdsourcing-Front, S. 2, 14. Juli 2011



### 3.3.6 Datenworkflow und -strategien entwickeln

Um als Datenjournalist erfolgreich zu sein, ist ein funktionierender Datenworkflow obligatorisch wichtig. Von der Aggregation der Daten bis zur fertigen Publikation sind Datenjournalisten aktiv am Entstehungsprozess beteiligt, je nach Umfang des Projekts entweder in einem interdisziplinären Team oder auch allein. Datenquellen, Bearbeitungswerkzeuge, Visualisierungstechniken, für die Umsetzung verfügbare Ressourcen, die Inhalte, die den Workflow ausmachen, sind vielfältig und richten sich ebenso nach der persönlichen Arbeitsweise, als auch Vorkenntnissen und den Möglichkeiten, die Redaktionen und Medien ihren Journalisten bereitstellen. Unterstützung für ihre Arbeit finden Datenjournalisten dabei vor allem in den Medien, die eine Strategie der Transparenz und Offenheit gegenüber ihren Lesern verfolgen. Denn das fördert das Vertrauen der Leser in die Medien und bietet Datenjournalisten die Möglichkeit Leser zu aktivieren und mittels Crowdsourcing an Datenprojekten teilhaben zu lassen und von deren multiperspektivischen Erkenntnissen in der Berichterstattung zu profitieren.

## 4. Datenworkflow für Journalisten

Informationen in Rohdaten zu finden und daraus verwertbares Wissen zu generieren ist die Kernaufgabe des Datenjournalismus. Um diese Aufgabe zu lösen, arbeiten Journalisten Teilprozesse ab, in denen Rohdaten gesammelt, aufbereitet, analysiert und die aus den Daten gewonnenen Informationen in den neuen Darstellungsformen des Datenjournalismus präsentiert werden. Datenjournalisten setzen dabei Techniken der explorativen Datenanalyse ein, die dem 1989 vorgestellten Prozessmodell des *Knowledge Discovery in Databases* (KDD) folgen und eine effektive Datenarbeit ermöglichen. KDD wird dabei als nicht-trivialer Prozess beschrieben, mit dem gültiges, neues und nutzvolles Wissen aus großen Datenmengen automatisiert extrahiert werden kann. Für die Extraktion nutzt KDD Methoden des Data Mining, die in Form von Algorithmen relevante Daten auswählen, vorverarbeiten und neu ausgeben<sup>154</sup>. Die so ersichtlichen Muster strukturierter Daten können anschließend visualisiert und analysiert werden und bilden so die Grundlage zur Extraktion von neuem Wissen aus den Daten.

Neue Anwendungsfelder, beispielsweise in der Bioinformatik oder Analyse von Klimadaten, reizten die Möglichkeiten der Data Mining-Methoden des KDD-Modells jedoch aus<sup>155</sup>. Denn mit der steigenden Komplexität der Daten fehlte es den Nutzern mehr und mehr an zeitsparenden und effizienten Möglichkeiten bereits während dem Analyseprozess die Daten zu visualisieren und die Algorithmen anpassen zu können. Die Bestrebungen nach einem flexibleren Prozessmodell führten zur Erweiterung der Data Mining-Methoden durch den Visual Analytics-Ansatz<sup>156</sup>. Dieser beschleunigt auf der einen Seite die Extraktion relevanter Informationen aus der Datenflut durch die Integration von Visualisierungstechniken in alle Ebenen des Analyseprozesse<sup>157</sup>. Auf der anderen Seite ersetzt Visual Analytics den Schritt des Data Mining im KDD-Modell mit einem Zyklus, der dem Nutzer die stetige Datenmanipulation und Ergebniskontrolle ermöglicht<sup>158</sup>. Dabei kombiniert Visual Analytics die Leistungsfähigkeit elektronischer Datenverarbeitungssysteme mit der menschlichen Fähigkeit visuelle Muster schnell erkennen und verarbeiten sowie sein eigenes Wissen in den Prozess mit einbringen zu können<sup>159</sup>.

Gegenüber dem weit verbreiteten KDD-Modell bietet Visual Analytics für Datenjournalisten folgende Vorteile bei der Analyse komplexer Datensätze:

- Einfache Kontrolle und schnelle, zyklische Anpassung der Data Mining-Methoden zur Extraktion von Mustern und Wissen aus den Daten.

---

154 Vgl. Fayyad et al. (1996): From Data Mining to Knowledge Discovery in Databases, AI Magazine 3/1996, S. 40 ff.

155 Vgl. Keim et al. (2010): Mastering the Information Age, S. 41

156 Vgl. ebenda, S. 3

157 Vgl. Keim et al. (2006): Challenges in Visual Data Analysis, S. 1

158 Vgl. Keim et al. (2010): Mastering the Information Age, S. 90

159 Vgl. ebenda, S. 2

- Flexible und agile Experimentierumgebung mit schneller Ergebniskontrolle durch die Integration interaktiver Datenvisualisierungen.
- Besondere Eignung für die Analyse und Visualisierung raumzeitlicher Daten, wie sie beispielsweise im Projekt *Verräterisches Handy* von ZEIT ONLINE angefallen sind (s. Kapitel 3.2.2)<sup>160</sup>.
- Visual Analytics eignet sich ebenso für die Mustererkennung in komplexen und mehrdimensionalen Datensätzen durch die Kombination interaktiver Visualisierungen mit Data Mining-Algorithmen<sup>161</sup>.

Die Bandbreite datenjournalistischer Projekte reicht von der Arbeit mit gut strukturierten Datensätzen in einfachen Visualisierungen bis zur Verarbeitung großer, mehrdimensionaler Datensätze in interaktiven Datenvisualisierungen. Datenjournalisten benötigen daher einen Workflow, der systematisch und trotzdem flexibel auf das breite Anforderungsfeld reagieren kann. Der Workflow soll Journalisten gleichzeitig helfen, in den einzelnen Arbeitsschritten nicht die Orientierung zu verlieren und verhindern, dass sie durch eine „[b]lind application of Data-mining methods [...] meaningless and invalid patterns“ entdecken<sup>162</sup>.

In diesem Kapitel wird deshalb ein Workflow für Datenjournalisten vorgestellt, der sich an den Möglichkeiten von Visual Analytics auf Basis des KDD-Modells orientiert und eine effektive und erfolgreiche Realisierung von Projekten auf Datenbasis sichern soll. Aufgrund der thematischen Komplexität und Informationsfülle der allgemeinen Datenanalyse, wird – um dem Rahmen dieser Arbeit zu entsprechen – in diesem Kapitel nur auf Prozesse eingegangen, deren Einsatz im datenjournalistischen Alltag eine wesentliche Rolle einnimmt. Damit ist dieses Kapitel auch ein Versuch, die Methoden des *Knowledge Discovery in Databases* für die speziellen Anforderungen des Datenjournalismus anzupassen und Handlungsempfehlungen auszusprechen, die ungeübten Journalisten den Einstieg in komplexere, datenjournalistische Projekte erleichtern sollen.

## 4.1 Start und Vorbereitung

Der Beginn einer Datenberichterstattung kann durch zwei Impulse ausgelöst werden: Der erste ist die Existenz einer konkreten Frage, die durch Daten am besten beantwortet werden kann<sup>163</sup>. Solche Fragen entstehen wie im klassischen Journalismus unter anderem durch eigene Beobachtungen, das Hinterfragen aktueller Meldungen und Ereignisse, Hinweise von Informanten oder Anregungen von außen<sup>164</sup>. Der zweite Impuls ist die Verfügbarkeit von Daten, die nach interessanten Mustern und Informationen durchsucht werden können. Die journalistische Story ergibt sich bei diesem Ansatz meist erst während des Analyse-

---

<sup>160</sup> Vgl. Keim et al. (2010): Mastering the Information Age, S. 42 f.

<sup>161</sup> Vgl. ebenda

<sup>162</sup> Vgl. Fayyad et al. (1996): From Data Mining to Knowledge Discovery in Databases, AI Magazine 3/1996, S. 39

<sup>163</sup> Vgl. Bradshaw (2010): How to be a data journalist, 15. Juli 2011

<sup>164</sup> Vgl. La Roche (1975): Einführung in den praktischen Journalismus, S. 53 f.

prozesses<sup>165</sup>. Davon ausgeschlossen sind Daten, die Journalisten durch Informanten oder Enthüllungsnetzwerke wie *Wikileaks* bereits mit einer entsprechenden Story zugespielt wurden. Vor einer Verwendung solcher Daten müssen diese allerdings gemäß der Sorgfaltspflicht aus Ziffer 2 des Pressekodex<sup>166</sup> erst auf ihre Echtheit geprüft werden.

### 4.1.1 Hintergrundrecherche

Nach der Initiierung eines Projekts müssen Journalisten erst einmal das thematische Umfeld und den Hintergrund eines Themas recherchieren, bevor sie in die Datenarbeit einsteigen können, denn Regierungsdaten zum Beispiel „*will often come littered with jargon and codes you'll need to understand*“<sup>167</sup>. Ebenso recherchieren Journalisten in der Startphase mögliche Quellen von Daten. Auch im KDD-Modell entspricht die Hintergrundrecherche dem ersten Arbeitsschritt und ist für eine erfolgreiche Datenanalyse wichtig. Denn das angeeignete Wissen hilft Datenjournalisten den Informationsgehalt der Daten einschätzen und Zusammenhänge zwischen den Daten besser verstehen zu können. Daher ist vor allem bei sehr komplexen Themen eine hinreichende Einarbeitung in das Thema notwendig.

So war im Fall der Fluglärmkarte der *Taz* zu Beginn des Projekts die Berechnung der individuellen Lautstärke eines jeden Flugzeugs in der finalen Anwendung geplant. Diese Funktion hätte das Projekt positiv erweitert, war „*[a]ber im Zeit- und Budgetrahmen [...] nicht realisierbar*“<sup>168</sup>. Aufgrund ihres Aufwands musste die Funktion zwar gestrichen werden, half den Projektteilnehmern aber auch, sich auf die umsetzbaren Kernfunktionen zu konzentrieren und keine direkte Arbeitszeit an Funktionen zu verschwenden, deren Entwicklung in Konflikt zu den vorhandenen Ressourcen steht.

### 4.1.2 Zielvorgaben definieren

Um die verfügbaren Ressourcen wie Zeit, Budget, Technik und Personal für ein Datenprojekt optimal auszureizen, ist es erforderlich sich zu Beginn realisierbare Ziele für ein Projekt zu setzen. Die Ziele hängen dabei von Art und Anzahl der verfügbaren Daten ab und können je nach Projekt variieren. Wichtige Fragen, die den Rahmen eines Projekts definieren, können folgende sein:

- Welche Frage(n) soll(en) durch die Daten beantwortet werden?
- Welche Datenquellen stehen zur Beantwortung der Frage(n) zur Verfügung?
- Welche Ergebnisse sind zu erwarten und wie sollen sie Lesern präsentiert werden?
- Welche technischen Anforderungen bringt das Projekt mit sich?
- Ist das für die Umsetzung notwendige Fachwissen vorhanden?
- Sollen den Lesern Werkzeuge zur Eigenrecherche bereit gestellt werden?

---

165 Vgl. Bradshaw (2010): How to be a data journalist, 15. Juli 2011

166 Vgl. Presserat (o.J.): Ziffer 2 - Sorgfalt, 14. Juli 2011

167 Bradshaw (2010): How to be a data journalist, 15. Juli 2011

168 Matzat (2011): Fluglärmkarte BBI: Making of, 14. Juli 2011

- Reicht das für die Umsetzung verfügbare Personal oder wird zusätzliches benötigt?
- Welche Hindernisse können auftreten und wie können sie vermieden werden?

Vor allem bei einem Projektstart über eine konkrete Frage kann es passieren, dass erst durch die Arbeit mit den Daten der Rahmen deutlich wird, in dem sich das Projekt bewegt. Denn durch die Datenanalyse können viele Anforderungen, Hindernisse oder Möglichkeiten aufgedeckt werden, die vorher nicht zu erkennen waren, und mehr Zeit als geplant in Anspruch nehmen. Für solche Veränderungen sollte der Projektplan möglichst flexibel gestaltet werden.

## 4.2 Datenaggregation

Nach der Festlegung der Projektziele beginnt das Sammeln der notwendigen Daten. Verfügen Journalisten bereits über Daten aus der Startphase, kann dieser Arbeitsschritt übersprungen werden. Während der Aggregationsphase sammeln Journalisten die Daten, welche es ihnen nach dem Ergebnis ihrer Recherche am wahrscheinlichsten ermöglichen werden Antworten auf ihre aufgestellten Fragen zu bekommen und passende Stories zu finden.

### 4.2.1 Datenbanken nutzen

Welche Quellen und Daten für ein Projekt genutzt werden, hängt primär von der Frage ab, die beantwortet und datenjournalistisch aufbereitet werden soll. So verarbeitet die Fluglärmkarte der *Taz* Prognosedaten zur Fluglärmbelastung des *Ministeriums für Infrastruktur und Landwirtschaft Brandenburg* (s. Kapitel 3.2.3), während das Projekt *Verräterisches Handy* von *ZEIT ONLINE* von der Telekom gespeicherte Vorratsdaten verarbeitet (s. Kapitel 3.2.2). Um an sinnvolle Daten zu gelangen, müssen Datenjournalisten wissen, welche Quellen über welche Daten verfügen. Neben den medieninternen Archiven, Informanten-Netzwerken oder Enthüllungsplattformen sind das vor allem Websites von Behörden oder Organisationen.

In der Auslandsberichterstattung oder im Wirtschaftsjournalismus können deutsche Medien die Websites *Data.gov* der US-amerikanischen und *data.gov.uk* der britischen Regierung als Datenquelle nutzen. Durch das Engagement der Open-Government-Bewegung (s. Kapitel 2.3.3) realisierten die Regierungen das Potential offener Daten und so stehen heute auf diesen Websites vielfältige Daten verschiedener Regierungsbehörden und -ministerien als Rohdaten zur Verfügung. Auf den Websites werden die Daten gebündelt und ohne Einschränkung frei zugänglich gemacht. Neben Dokumenten und Tabellen enthalten die Plattformen auch Geodaten, Links zu anderen Open-Data-Websites und Tipps zur Verarbeitung der Daten. Die Datensätze der Open-Government-Websites bieten hilfreiche Daten unter anderem aus dem Industrie-, Landwirtschafts-, Energie-, Gesundheits- und Umweltsektor. Zwar befinden sich ähnliche Open-Government-Portale in Deutschland noch im Aufbau, trotzdem existieren auch zur Inlandsberichterstattungen bereits Daten, die Journalisten

in Deutschland nutzen können. Das von Bund und Ländern aufgebaute *GENESIS*-Datenbanksystem, in dem die verfügbaren Daten der *Statistischen Landesämter* und des *Statistischen Bundesamts* auf einer Plattform zusammenlaufen, ist die umfangreichste öffentliche Datenbank zu aktuellen Daten über Deutschland. Das Datenbanksystem verfügt dabei über eine gute Suchfunktion, eine funktionale Filtermöglichkeiten zur Eingrenzung der Ausgabedaten sowie die Option die ausgegebenen Daten zur einfachen Weiterverarbeitung in MS-Excel-, CSV- oder HTML-Dateien zu speichern.

Auf internationaler Ebene stehen viele Datenquellen zur Verfügung. Für Daten der EU können Journalisten die Datenbanken des *Statistischen Amts der Europäischen Union* (EuroStat) nutzen, das bereitgestellte Daten aus den statistischen Ämtern seiner Mitgliedsstaaten aufbereitet und zur Verfügung stellt. Für weltweite Daten zu den Themen Bildung und Einkommen eignen sich die Datenbanken der *Vereinten Nationen*, Wirtschafts-Daten finden sich in den Datenbanken der *Welthandelsorganisation* und bei Datenprojekten zu Gesundheitsthemen bieten die Datenbestände der *Weltgesundheitsorganisation* eine hohe Informationsfülle an.

## 4.2.2 Daten selbst aggregieren

Sind keine spezifischen Daten in Datenbanken vorhanden, bleibt Datenjournalisten nur die Möglichkeit selbst Daten zu sammeln, auf deren Grundlage sie ihre Berichterstattung aufbauen können. Abhängig von der Art der Daten, die gesammelt werden sollen, gibt es verschiedene Methoden selbst Daten zu aggregieren.

### 4.2.2.1 Archivmaterial nutzen

Bei dem Projekt *Todesopfer rechter Gewalt* von *ZEIT ONLINE* wurde investigativ die Anzahl von Tötungsdelikten durch Rechtsextremismus zwischen 1990 und 2010 untersucht und die Ergebnisse in einer interaktiven Grafik visualisiert<sup>169</sup>. Dazu sichtete *ZEIT ONLINE* „[h]underte Lokalzeitungsartikel und Gerichtsurteile“<sup>170</sup>, interviewte „Opferberater, Hinterbliebene, Anwälte und Strafverfolger“<sup>171</sup> und deckte durch die intensive Recherchearbeit eine fast dreimal höhere Zahl von Todesopfern rechter Gewalt auf, als offizielle Statistiken verzeichneten<sup>172</sup>.

### 4.2.2.2 Screen Scraping

Eine weitere Methode zur Datenaggregation ist das Screen Scraping. Dabei werden die auf Websites in Textform gelagerten Informationen durch Programme, sogenannte „Screen Scraper“, ausgelesen und in einer Tabellen-Datei gespeichert, um die Daten einfacher sortieren, filtern und weiter verarbeiten zu können. Der Aufwand, um die Daten auszulesen, ist technisch jedoch anspruchsvoll und erfordert einige Programmierkenntnisse zur Erstellung der Screen Scraper. Haben Medienhäuser bereits den Schritt zum Aufbau von

169 ZEIT ONLINE (2010): Todesopfer rechter Gewalt 1990 - 2010, 15. Juli 2011

170 Langner (o.J.): Thema Datenjournalismus: Das Beispiel „Zeit online“, 15. Juli 2011

171 ebenda

172 Vgl. Jansen et al. (2010): Eine furchtbare Bilanz, 15. Juli 2011



eigenen Datenteams gewagt oder verfügen über Zugang zu externen Dienstleistern mit entsprechendem Wissen, ist die Extraktion von Daten aus Websites über Screen Scraping für Datenjournalisten eine exklusive Informationsquelle.

Die Website *scraperwiki.com* bietet jedoch den kostenlosen Service Anfragen zur Aggregation von Daten bei fehlendem technischem Verständnis an die eigene Programmierer-Community weiterzuleiten<sup>173</sup>. So können auch Datenjournalisten ohne Vorerfahrungen an Daten gelangen, datengestützte Berichterstattungen erarbeiten und ihre Kompetenzen im Datenjournalismus ausbauen. Außerdem sind auf der Website bereits fertig programmierte Scraper kostenlos verfügbar, die Datenjournalisten mit Programmiergrundkenntnissen für ihre Bedürfnisse anpassen können.

Neben den selbst programmierten Eigenentwicklungen lassen sich Daten aber ebenso über kostenlose Programme wie *Google Spreadsheets* oder *Yahoo Pipes* scrapen, deren Bedienung keine besonderen Programmierkenntnisse voraussetzt und im Internet gut dokumentiert und erklärt ist.

Datenjournalisten sollten allerdings beachten, dass sich der Einsatz von Screen Scraping in einer rechtlichen Grauzone bewegt, da Urheberrechtsfragen für das automatische Auslesen von im Internet veröffentlichten Daten, noch nicht geklärt sind<sup>174</sup>. Nicht-öffentliche Bereiche von Websites wie zum Beispiel geschlossene Foren, zu denen der Zugang nur über eine Nutzerregistrierung möglich ist und in denen eine Vielzahl von Nutzern private Informationen hinterlassen, sollten von Datenjournalisten generell vom Scraping ausgeschlossen werden.

## 4.3 Datenaufbereitung

Aufgrund der vielfältigen Quellen, die Datenjournalisten zur Verfügung stehen, kann die Qualität und Konsistenz der Daten nach der Aggregationsphase stark variieren<sup>175</sup>. Um die Daten effektiv weiter verarbeiten zu können, ist es notwendig ihre Qualität zu prüfen und bestehende Mängel auszubessern. Deswegen beschäftigt sich der vierte Arbeitsschritt im Datenworkflow mit der Aufbereitung von Daten und hat zum Ziel, die Daten für die optimale Weiterverarbeitung in geeignete Dateiformate umzuwandeln und von Fehlern zu säubern.

### 4.3.1 Datenumwandlung

#### 4.3.1.1 Daten aus Ausdrucken extrahieren

Besteht die gesammelte Datenbasis eines Projekts aus ausgedruckten Dokumenten in Textform, ist es erforderlich die Daten zu digitalisieren. Hierzu werden Programme eingesetzt, die mittels *Optical Character Recognition*-Verfahren (OCR) den Text automatisch erkennen und wahlweise in Textdateien oder Tabellen umwandeln. Textdateien werden benötigt, wenn der Text an sich – wie im Fall der Doktorarbeit von Karl-Theodor zu Guttenberg (s. Kapitel 3.2.5) – die Datenbasis bildet und in seiner Ganzheit erhalten bleiben soll.

---

173 Vgl. Scraperwiki (o.J): Request data, <http://bit.ly/scraperwikirequest>, 20. Juli 2011

174 Vgl. Olorode (2010): Scraping als neue Methode zur Erhebung persönlicher Daten, 15. Juli 2011

175 Vgl. Keim et al. (2010): Mastering the Information Age: Solving problems with visual analytics S. 26

Enthalten die Dokumente dagegen Daten in Form von Zahlen, ist eine Ausgabe als Tabelle zur Weiterverarbeitung sinnvoll.

Um OCR-Programme einsetzen zu können, müssen die ausgedruckten Dokumente erst eingescannt und als Bilder gespeichert werden. Danach analysiert die Software die Buchstaben, erkennt diese und extrahiert die enthaltenen Daten in das entsprechende Dateiformat. Eine hohe Schriftqualität der eingescannten Dokumente erhöht dabei die Genauigkeit des Verfahrens. Besonders bei noch nicht digitalisiertem Archivmaterial und sehr großen Datensätzen erleichtert der Einsatz von OCR-Software Datenjournalisten die weitere Arbeit. Neben kommerziellen Angeboten können Journalisten auch die OCR-Funktion von *Google Docs* als kostenlose Alternative verwenden.

#### 4.3.1.2 Daten aus Bildern extrahieren

Das OCR-Verfahren kann auch angewendet werden, um Daten aus Texten zu extrahieren, die als Bilder gespeichert wurden. So ein Einsatz ist bei Texten möglich, die durch falsches Dateimanagement in reinen Bilddateien oder in PDF-Dateien, die aus Bildern generiert wurden, vorliegen.

#### 4.3.1.3 Daten aus Tabellen in PDF-Dateien extrahieren

Daten, die als Tabelle in einer PDF-Datei vorliegen, sind für eine Weiterverarbeitung nicht geeignet und müssen in eine Tabellendatei umgewandelt werden. Denn Tabellen einer PDF können nur zeilenweise oder zellenweise selektiert werden, wodurch beispielsweise Daten einer Spalte nur schrittweise kopiert werden können. Bei großen Tabellen ist dieser Prozess langwierig und zeitraubend. Mithilfe der Programme *PDF2XL* oder *cometdocs*, die auch von der *Deutschen Presse Agentur* (dpa) eingesetzt werden<sup>176</sup>, lassen sich Daten aus Tabellen in PDF-Dateien jedoch schnell und einfach auswählen und speichern.

### 4.3.2 Datenbereinigung

Durch solche Umwandlungsprozesse, aber auch durch die ungenaue Arbeit von Regierungsbehörden oder das Sammeln von eigenen Daten können Fehler in den Datensätzen entstehen. Durch die Datenbereinigung wird sichergestellt, dass keine selbst verursachten Anomalien in der Datenmasse auftauchen, die in der späteren Analysephase zu einer falschen Interpretation der Daten durch den Analysten führen können<sup>177</sup>. Das würde nicht nur die komplette Berichterstattung verfälschen, sondern auch die journalistische Glaubwürdigkeit des Mediums verringern, das seine Publikationen auf den Daten aufbaut.

Folgende Fehler können in Daten auftreten<sup>178</sup>:

- Unvollständige Daten durch Aggregations- oder Umwandlungsfehler

---

<sup>176</sup> Vgl. Elmert (2011): Datenjournalismus ganz praktisch, S. 11 f.

<sup>177</sup> Vgl. Keim et al. (2008): Visual Analytics: Definition, Process, and Challenges, Information Visualization, S. 167

<sup>178</sup> Vgl. Keim et al. (2010): Mastering the Information Age: Solving problems with visual analytics, S. 26



- Inkonsistente Daten durch variierende Formatierungen, Abkürzungen oder Rechtschreibfehler
- Falsche Zahlenwerte durch Rechen-, Rechtschreib- oder Umwandlungsfehler

Für die Bereinigung eignen sich als Werkzeuge zum Beispiel *Microsoft Excel* oder *Google Refine*. Letzteres ist empfehlenswert, da es kostenlos angeboten wird, gut mit anderen *Google* Produkten zusammenarbeitet, übersichtlich aufgebaut ist und über zeitsparende Funktionen für die Datenbereinigung verfügt.

## 4.4 Visuelle Datenanalyse

Die Analyse der aggregierten Daten ist der wichtigste Schritt im Datenjournalismus und der Kern einer datengestützten Berichterstattung, da hier die Entdeckung von neuem Wissen stattfindet. Gleichzeitig ist er auch der Arbeitsschritt, der die meiste Arbeit verursacht<sup>179</sup>. Durch die hohe Flexibilität und den schnellen Erkenntnisgewinn des Visual Analytics-Modells ist es Datenjournalisten möglich, große und komplexe Datenstrukturen in kurzer Zeit zu durchdringen, Muster und Zusammenhänge in den Daten zu entdecken und Wissen aus den Daten zu extrahieren.

Visual Analytics kombiniert automatisierte Analyseverfahren mit interaktiven Visualisierungen, um ein „*effective understanding, reasoning and decision making*“<sup>180</sup> auf Basis komplexer Datensätze zu erreichen. Aufbauend auf dem von dem deutschen Informatiker Daniel Keim entwickelten Mantra „*Analyse first, show the important, zoom, filter and and analyse further, details on demand*“<sup>181</sup>, ist der Prozess der Wissensextraktion über Visual Analytics ein Kreislauf aus Visualisierung, Analyse und der Anpassung von falschen oder ungenauen Parametern und Filtern, bis alle in den Daten enthaltenen Informationen aufgedeckt worden sind.

Die Ziele dieser visuellen Datenanalyse<sup>182</sup> sind

- Informationen aus Daten synthetisieren und daraus Erkenntnisse zur Beantwortung der Hypothese ableiten
- Das Erwartete erfassen und Unerwartetes entdecken
- In kurzer Zeit vertretbare und verständliche Feststellungen erarbeiten
- Feststellungen als Grundlage für weitere Handlungen kommunizieren

---

179 Vgl. Fayyad et al. (1996): From Data Mining to Knowledge Discovery in Databases, *AI Magazine* 3/1996, S. 42

180 Keim et al. (2008): Visual Analytics: Definition, Process, and Challenges, *Information Visualization*, S. 157

181 Keim (2005): Scaling Visual Analytics to Very Large Data Sets, S. 33

182 Vgl. Keim et al. (2008): Visual Analytics: Definition, Process, and Challenges, *Information Visualization*, S. 167

### 4.4.1 Visualisierungstechniken

Als erster Schritt der Analyse müssen die Daten visualisiert werden. Dazu stehen verschiedene Techniken bereit. Nach Daniel Keim werden die für die visuelle Analyse verfügbaren Techniken nach den Kriterien des Datentyps, der Visualisierungstechnik und der Interaktions- und Verzerrungstechnik unterschieden, die die folgende Klassifikation ergeben:

Datentyp	Visualisierungstechnik	Interaktions- und Verzerrungstechnik
Ein-dimensionale Daten	Standard 2D-/3D-Darstellung	Standard
Zwei-dimensionale Daten	Geometrische Transformation	Projektion
Multi-dimensionale Daten	Icon-basierte Darstellung	Filterung
Text und Hypertext	Pixel-Darstellung	Zoom
Algorithmen und Software	Link und Brush	

**Tabelle 1:** Klassifikation der Techniken zur visuelle Datenanalyse<sup>183</sup>

Keim weist zudem daraufhin, „*that the three dimensions of our classification [...] can be assumed to be orthogonal*“<sup>184</sup>. Danach kann also auf jeden Datentyp jede Visualisierungs-, Interaktions- oder Verzerrungstechnik bzw. sogar eine Kombination verschiedener Techniken angewendet werden<sup>185</sup>.

An dieser Stelle kann der hier beschriebene Workflow keine genaue Anleitung zur Auswahl von Techniken ausgeben, da – neben Art und Umfang der Rohdaten selbst – auch die für ein Projekt verfügbaren Ressourcen wie Zeit, Budget, Technik und Personal in die Entscheidung mit einfließen und sich projektweise zu stark unterscheiden. Die gewählte Technik oder Kombination von Techniken sollte im Idealfall aber auf alle der folgenden Punkte zutreffen, die als Orientierungshilfe für Datenjournalisten dienen sollen:

- Vereinfacht die Visualisierung die Datenstruktur und erleichtert die Analyse und Interpretation der Daten?
- Bietet die Visualisierung einen Überblick über alle Daten?
- Würden zusätzliche Visualisierungen die Übersichtlichkeit verbessern?
- Können durch die Visualisierung Details in den Daten ausgelesen werden?
- Ist bei großen Datenmassen eine Filterung durch Interaktionstechniken möglich?
- Werden Muster und Zusammenhänge zwischen den Daten deutlich?
- Lässt sich die Visualisierung schnell anpassen und mit neuen Daten ergänzen?
- Ist das Projekt mit der Visualisierung im Rahmen der verfügbaren Ressourcen umsetzbar?

### 4.4.2 Muster erkennen und interpretieren

Nach der Auswahl der Visualisierungstechnik und Anwendung auf die Rohdaten können

<sup>183</sup> Vgl. Keim (2002): Information Visualization and Visual Data Mining, S. 101

<sup>184</sup> ebenda

<sup>185</sup> Vgl. ebenda

sich Datenjournalisten im Sinne des Wortes ein erstes Bild von den Daten machen. Ausgehend von der Fragestellung aus der Startphase werden die Daten im weiteren Verlauf auf allgemeine Auffälligkeiten untersucht, was zur Bildung erster Hypothesen führen kann. Durch eine intensivere Datenarbeit und Suche nach erklärenden Muster und Zusammenhänge müssen die Hypothesen anschließend belegt werden.

Gleich der Empfehlung, die Walther von La Roche bei der allgemeinen Recherche im klassischen Journalismus ausspricht<sup>186</sup>, sollten Datenjournalisten auch bei der Untersuchung von Datenvisualisierungen von außen nach innen vorgehen. Dadurch verschaffen sie sich erst einen allgemeinen Überblick und inspizieren den gesamten Datenbestand, um anschließend mit einem besseren Verständnis tiefer in die Daten vorzudringen und Muster und interessante Details aufdecken zu können. Nach Keims Zielen der visuellen Datenanalyse kann die Untersuchung des Gesamtbildes als Erfassung der Erwartungshaltung und die Detailarbeit als Entdeckung des Unbekannten gesehen werden.

Ist der Datensatz im Umfang sehr komplex und lässt sich auch nicht in angemessener Zeit durch einzelne Datenjournalisten oder ein verfügbares Team analysieren, ist es erforderlich, dass Medien frühzeitig Strukturen und Plattformen aufstellen, um an diesem Punkt die Öffentlichkeit in den Prozess der Datenanalyse mit einbinden und die Aufdeckung aller Informationen beschleunigen zu können.

Anhand der durch Crowdsourcing oder die Arbeit Weniger gewonnenen Informationen können anschließend die aufgestellten Hypothesen auf ihren Wahrheitsgehalt untersucht und mit den daraus resultierenden Ergebnissen die anfängliche Frage beantwortet werden. In diesem Wechselspiel der verschiedenen Schritte und Prozesse der Datenanalyse findet die Extraktion von Wissen aus Daten statt und bildet den Stoff für die darauf aufbauende journalistische Berichterstattung.

### 4.4.3 Anwendungsgebiete

Obgleich die visuelle Datenanalyse in vielen Bereichen mit hohem Datenaufkommen eingesetzt werden kann, ist ihr Einsatz im Datenjournalismus besonders bei Wirtschafts-, sozioökonomischen und ortsgebundenen Daten sinnvoll.

Da im Wirtschaftsjournalismus täglich über Wirtschaftsdaten, wie Kursveränderungen in den Aktienmärkten und aktuelle Geschäftsberichte, berichtet wird, sind Journalisten bei ihrer Berichterstattung einem Strom heterogenen Informationen aus verschiedenen Quellen in hoher Abfolge ausgesetzt<sup>187</sup>. Der Einsatz visueller Datenanalyse-Techniken in diesem Gebiet ist daher ein nützliches Werkzeug, damit Datenjournalisten die komplexen Zusammenhänge wirtschaftlicher Entwicklungen verstehen und darüber berichten können.

Sozioökonomische Daten liefern Journalisten Informationen, um gesellschaftliche Prozesse im Zusammenspiel mit wirtschaftlichen Entwicklungen zu beschreiben, beispielsweise der Zusammenhang zwischen dem Wahlverhalten einer Region, deren Demografie und Wirtschaftskraft. Medien als Instrument der öffentlichen Meinungsbildung profitieren von einem schnellen Zugang zu diesen Informationen in den Daten über das Mittel der visuellen

---

<sup>186</sup> Vgl. La Roche (1975): Einführung in den praktischen Journalismus, S. 59

<sup>187</sup> Vgl. Keim et al. (2008): Visual Analytics: Definition, Process, and Challenges, Information Visualization, S. 165 f.

Datenanalyse und sind dadurch auch in der Lage die Wechselbeziehungen einer modernen Gesellschaft zwischen Wirtschaft, Kultur und Politik zu verstehen<sup>188</sup>. Somit ist der Einsatz von Techniken der visuellen Datenanalyse neben dem Wirtschaftsjournalismus auch für den datengestützten Politik- und Kulturjournalismus sinnvoll, wie auch am Beispiel des Projekts *Verräterisches Handy* von ZEIT ONLINE deutlich wird (s. Kapitel 3.2.2).

Das Potential ortsgebundener Daten durch kartografische Visualisierungstechniken für den Lokaljournalismus wurde bereits am Beispiel der Fluglärmkarte der *Taz* in Kapitel 3.2.3 beschrieben.

## 4.5 Datenpräsentation

*„The ultimate goal is to bring the power of visualization technology to every desktop to allow a better, faster and more intuitive exploration of very large data resources. This will not only be valuable in an economic sense but will also stimulate and delight the user“*

(Daniel Keim, deutscher Informatiker)<sup>189</sup>

Diese Vision von Daniel Keim aus dem Jahr 2002, dass mithilfe von Methoden und Techniken der visuellen Datenanalyse es jedem möglich sein wird, komplexe Datenmassen durchdringen und verstehen zu können, kann durch Datenjournalismus Wirklichkeit werden. Denn neben dem Potential von Datenjournalismus für die Medien (s. Kapitel 3) sind interaktive Datenvisualisierungen und besonders datengestützte Crowdsourcing-Projekte eine Möglichkeit Nutzer für Daten zu sensibilisieren und selbst in der Datenanalyse aktiv werden zu lassen (s. Kapitel 3.2.1).

Um Leser und Nutzer aktivieren und ihnen das aus den Daten gewonnene Wissen so übersichtlich und zugleich so umfangreich wie möglich darlegen zu können, sollten Datenjournalisten bei der Präsentation eines datengestützten Projekts folgende Punkte beachten, die sich aus der Analyse vorhandener Beispiele im Datenjournalismus ableiten:

### 4.5.1 Werkzeuge zur Interaktion und Kollaboration integrieren

Wie die Beispiele in Kapitel 3 zeigen, sind die im Datenjournalismus verarbeiteten Daten meist sehr groß und enthalten eine Fülle an Informationen. Einzelne statische Datenvisualisierungen könnten diese Fülle nicht darstellen, weswegen bereits die Aufbereitung und Analyse der Daten durch interaktive Datenvisualisierungen nach dem Ansatz von Visual Analytics vorgenommen wird. Dementsprechend müssen bei der Präsentation komplexer Daten in Medien ebenfalls interaktive Datenvisualisierungen genutzt werden, die *„intelligently combine visualizations of selected analysis details on the one hand, and a global overview on the other hand“*<sup>190</sup> und es Nutzern ermöglichen nicht nur eine festgelegte Perspektive,

188 Vgl. Keim et al. (2008): Visual Analytics: Definition, Process, and Challenges, Information Visualization, S. 165 f.

189 Keim (2002): Information Visualization and Visual Data Mining, S. 105

190 Keim et al. (2008): Visual Analytics: Definition, Process, and Challenges, Information Visualization, S. 167

sondern sowohl die Gesamtheit der Daten, als auch deren Details wahrnehmen zu können. Durch den Einsatz interaktiver Werkzeuge zur Selektion und Filterung, wie eine Suchfunktion oder vordefinierte Kategorien von Daten, können Nutzer die Daten selbst analysieren, sich aktiv mit ihnen auseinandersetzen und Details in den Daten entdecken.

Ist eine Beteiligung der Öffentlichkeit bei der Analyse der Daten durch Crowdsourcing bei den Medien gewünscht, sollte zudem eine Kommunikationslösung zwischen dem Medium und seinen Lesern in der Datenanwendung integriert werden, um mögliche Entdeckungen der Leser direkt empfangen zu können und in die Berichterstattung mit einbauen zu können. Die Integration solcher Werkzeuge in interaktive Datenvisualisierungen erfordert unter Umständen einigen Programmieraufwand, aber der Erfolg von datenjournalistischen Projekten wie dem Spesenskandal des Guardian belegt den Nutzen der Werkzeuge nicht nur für die Medien.

#### 4.5.2 Aufbereitete Rohdaten ohne Einschränkung bereitstellen

Zum Zeitpunkt der Veröffentlichung des Projekts *Verräterisches Handy* waren die Informationen der dem Projekt zugrunde liegenden Daten bereits durch das von *OpenDataCity* und *ZEIT ONLINE* gewählte Visualisierungsverfahren komplett erschlossen (s. Kapitel 3.2.2). Obwohl eine Auslagerung der Analyse durch Crowdsourcing nicht nötig war, stellte *ZEIT ONLINE* die Daten seinen Lesern uneingeschränkt zur Verfügung. Neben der in Kapitel 3.3.4 aufgeführten Möglichkeit zum Aufbau von Vertrauen und Transparenz, ermöglicht eine Bereitstellung der Rohdaten auch Nutzern eigene Recherchen durch die Daten einzuleiten und auf der Arbeit der Medien aufzubauen. Möglich wird dadurch ebenso, dass Nutzer mit ihrem Wissen die gewählte Visualisierungstechnik verfeinern und bisher nicht ersichtliche Zusammenhänge in den Daten aufdecken zum Beispiel durch die Fusion mit anderen Datensätzen. Demnach sollten Datenjournalisten, die in ihren Projekten nicht auf die Arbeit der Öffentlichkeit angewiesen sind, ihre aufbereiteten Rohdaten trotzdem veröffentlichen, um langfristige Synergien zwischen Medien und Öffentlichkeit nicht auszuschließen.

Bei der Veröffentlichung von Daten sollten sich Datenjournalisten an dem von Tim Berners-Lee auf der *Gov 2.0 Expo 2010* vorgestellten 5-Sterne-System zur Bewertung der „Offenheit“ von Daten<sup>191</sup> orientieren:

- 1 Stern:** Daten generell online veröffentlichen
- 2 Sterne:** Strukturierte Daten veröffentlichen (z.B. als CSV- oder Excel-Datei, statt einem eingescannten Bild einer Tabelle)
- 3 Sterne:** Verwendung eines nicht-proprietären Datenformats (z.B. CSV statt Excel)
- 4 Sterne:** URLs benutzen, damit auf bestimmte Daten und Objekte genau verlinkt werden kann
- 5 Sterne:** In den eigenen Daten zu den Daten anderer verlinken, um einen größeren datengestützten Kontext anzubieten

---

191 Vgl. Berners-Lee (2010): (Video, ab 0:00 bis 2:00)

### 4.5.3 Interessante Details einzeln aufbereiten

Am Beispiel der Berichterstattung der New York Times über die afghanischen Kriegstagebücher (s. Kapitel 3.1.6) zeigt sich, dass es für Medien vorteilhaft ist interessante Details, die in den Daten gefundenen wurden, in klassischen Einzelbeiträgen journalistisch zu verarbeiten, um über ein Thema intensiv berichten zu können. So finden Leser auf der für die Berichterstattung angelegten Übersichtsseite nicht nur eine Gesamteinschätzung des Krieges, sondern auch ausgewählte Dokumente<sup>192</sup>, die eine eigene Geschichte erzählen und die Wahrnehmung über das behandelte Thema schärfen. Ohne die gesonderte Herausstellung, würden solche Einblicke in der Masse der präsentierten Daten verloren gehen. Datenjournalisten sollten daher eng mit ihren Kollegen im Print- und Online-Journalismus zusammenarbeiten und auch die Details und kleinen Geschichten der Daten in die allgemeine Berichterstattung mit einbauen, die die Qualität einer datenjournalistischer Berichterstattungen erhöhen kann.

### 4.5.4 Richtlinien guter Datenvisualisierungen einhalten

Die Datenvisualisierung ist das im Datenjournalismus wichtigste Instrument, um einen schnellen Wissenstransfer zwischen Medien und Lesern durchzuführen. Datenjournalisten sollten sich daher bei der Konzeption ihrer interaktiven Datenanwendungen an die Grundsätze guter Datenvisualisierungen bzw. Informationsvisualisierungen halten, die Regeln aufstellen, nach welchen Kriterien Informationen darzustellen sind, um eine optimale Informationsaufnahme zu gewährleisten.

---

192 Vgl. o.V. (2010): View Is Bleaker Than Official Portrayal of War in Afghanistan, 15. Juli 2011

## 5. Datenvisualisierungen

Mithilfe computergestützter Technologien kann heute jeder in kurzer Zeit Datenvisualisierungen erstellen. Nur leider fehlt es vielen an Verständnis dafür, was eine gute Visualisierung ausmacht<sup>193</sup>. Vielfach wird versucht die Aufmerksamkeit von Rezipienten auf eine Visualisierung durch unnütze Zusätze wie 3D-Balken, intensivem Farbeinsatz oder grafischen Elementen zu erregen<sup>194</sup>. Dabei wird der eigentliche Sinn der Visualisierung, Informationen aufzuzeigen und weiterzugeben, verfälscht und die *„presentation fails to convey the intended message“*<sup>195</sup>.

Damit Datenjournalisten bei der Konzeption und Präsentation ihrer Datenvisualisierungen solchen Fehlern entgehen und ihre Daten und das daraus gewonnene Wissen aussagekräftig und prägnant visualisieren können, ist es für sie wichtig zu wissen, wie Visualisierungen beschaffen sein müssen, um eine optimale Informationsaufnahme zu garantieren. Deswegen setzt sich dieses Kapitel mit den Grundsätzen guter Datenvisualisierungen auseinander und fasst die Erkenntnisse anhand eines aktuellen Beispiels aus dem Datenjournalismus zusammen.

Das verwandte Themenfeld der Informationsgrafiken, die in ihrer Funktion mehr Sachverhalte und Prozesse erklären anstatt neues Wissen aus Daten zu extrahieren, findet dagegen in diesem Kapitel keine Beachtung.

### 5.1 Grundsätze guter Datenvisualisierungen

*„Graphical elegance is often found in simplicity of design and complexety of data.“*

(Edward R. Tufte, US-amerikanischer Informationswissenschaftler und Grafikdesigner)<sup>196</sup>

Edward Tufte, dessen Arbeit das Feld der Statistik und des Informationsdesign wesentlich beeinflusste, sieht den Kern einer guten Visualisierung in dem Zusammenspiel von Einfachheit in ihrer Gestaltung und der Komplexität der visualisierten Daten. Die Grenzen, in denen sich Datenvisualisierungen dabei bewegen sollten, sind einerseits die Reduktion komplexer Daten auf ihre wesentliche Information und das Wissen, das sie vermitteln sollen. Andererseits sollten interessante Details der Daten in der Visualisierung nicht untergehen. Das entspricht auch den Forderungen von Daniel Keim bei der Auswahl von Visualisierungstechniken (s. Kapitel 4.4.2).

---

193 Vgl. Wong (2010): The Wall Street Journal Guide to Information Graphics, S. 13

194 Vgl. Tufte (1983): The Visual Display of Quantitative Information, S. 61 ff.

195 Wong (2010): The Wall Street Journal Guide to Information Graphics, S. 14

196 Tufte (1983): The Visual Display of Quantitative Information, S. 177



### Attraktive Datenvisualisierungen nach Tufte<sup>197</sup>

- kombinieren Wörter, Nummern und Zeichnungen
- erzählen oft eine Geschichte über die Daten
- stellen die erreichbare Komplexität der Details dar
- besitzen ein geeigneten Formats und eine geeignete Gestaltung
- sind professionell und sorgfältig angefertigt
- spiegeln Daten in ihrer Skalierung ausgeglichen und proportional wider
- vermeiden irrelevante Ausschmückungen, die keinen Bezug zum Inhalt aufweisen.

Dem ist hinzuzufügen, dass die drei obersten Aussagen je nach Beschaffenheit der Daten und Zielvorgaben nicht immer auf eine Visualisierung zutreffen müssen, während die unteren vier Aussagen allgemeine Gültigkeit besitzen, denen jede Datenvisualisierung folgen sollte. Darüber hinaus sind folgende Grundsätze ebenfalls bei jeder Visualisierung zu berücksichtigen:

#### 5.1.1 Data-Ink

Eine Datenvisualisierung besteht aus zwei Teilen. Der erste Teil enthält die Daten der Visualisierung, der zweite Teil kann als das Gerüst einer Visualisierung angesehen werden, also den Skalen, Beschriftungen und Hinweisen, die zum Auslesen und zur korrekten Interpretation und Zuordnung der Daten in den Kontext nötig sind. Nach Tufte sollte eine Datenvisualisierung die Aufmerksamkeit des Betrachters auf nichts anderes lenken als auf die Daten selbst<sup>198</sup>. Als Bewertungsmaßstab von Visualisierungen führt er das Konzept der *Data-ink ratio* ein. Die *Data-ink ratio* ist ein Quotient, der aus der vorhandenen Fläche, die die Daten einnehmen, und der Gesamtfläche der Grafik berechnet wird<sup>199</sup> und angibt, wie hoch der Anteil der *Data-ink* und der *non-Data-ink* in einer Visualisierung ist. Umso näher sich die *Data-ink ratio* an 1,0 (100 Prozent) befindet, umso höher ist die Datendichte und umso stärker fokussiert sich die Visualisierung auf die Daten (s. Abb. 9). Anzumerken ist, dass das Konzept als grobe Richtlinie zur Beschaffenheit einer Datenvisualisierung gesehen werden sollte, da eine genaue Berechnung durch die subjektive Schätzung der Bestandteile nicht möglich ist und je nach Betrachter variieren kann. Gleichzeitig trifft das Konzept keine Aussage über eine ideale Datenvisualisierung, da je nach Art der Daten und der Visualisierungstechnik mehr oder weniger *non-Data-ink* notwendig ist, um die Visualisierung interpretieren zu können. Das Konzept setzt sich vielmehr damit auseinander, welche Elemente in Visualisierungen redundant sind und wie durch die Reduktion von Redundanzen die *Data-ink-ratio* von Datenvisualisierungen gesteigert werden kann, um dem Betrachter der Visualisierung eine höhere Daten- und Informationsdichte zu vermitteln<sup>200</sup>.

---

197 Vgl. Tufte (1983): The Visual Display of Quantitative Information, S. 177

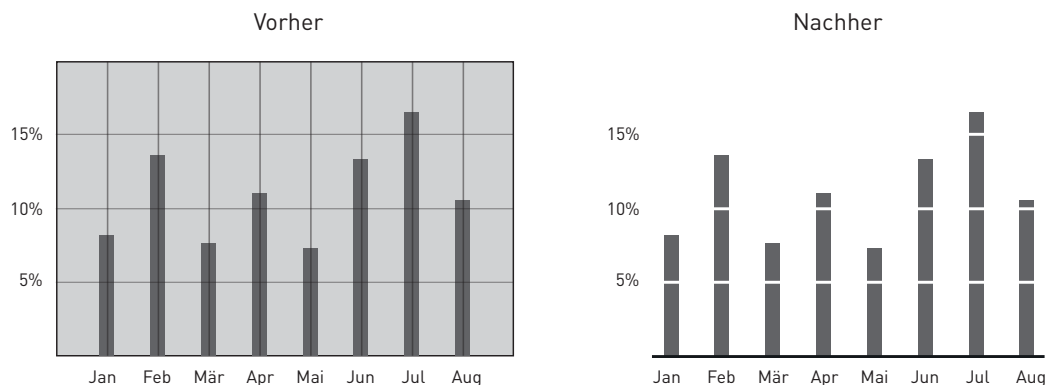
198 Vgl. ebenda, S. 91

199 Vgl. ebenda, S. 93

200 Vgl. Tufte (1983): The Visual Display of Quantitative Information, S. 96 ff.



Tuftes Ansatz folgend, dass jedes Stück „Tinte“ auf einer Grafik einen Grund braucht<sup>201</sup>, sollten Datenjournalisten bei der Gestaltung ihrer Visualisierungen darauf achten keine überflüssigen Elemente aufzunehmen, die Nutzer bei der Interpretation der Daten und Informationsaufnahme stören könnten.



**Abbildung 9:**

Erhöhung der *Data-ink ratio* eines Diagramms durch Reduktion

**Quelle:**

Nachbau nach Tufte

## 5.1.2 Lie Factor und Dimensionierung

Bei der Visualisierung von Daten ist es nicht nur wichtig die Daten in einer adäquaten Form zu visualisieren, sondern die Visualisierung so vorzunehmen, dass sie optisch auch den Daten entspricht. Tufte spricht in dem Zusammenhang von „*visueller Repräsentation*“. Diese wird erreicht, wenn die gemessene Größe visualisierter Daten ihrer numerischen Quantität entspricht<sup>202</sup>. Um visuelle Repräsentation von Datenvisualisierungen zu messen, führt Tufte den *Lie Factor* als Quotient aus der Größe der visuellen Darstellung von Daten und der Größe des numerischen Wertes von Daten ein<sup>203</sup>. Bei einer exakten Repräsentation beträgt der *Lie Factor* 1,0 und umso größer er wird, desto weniger entspricht die Visualisierung den ihr zugrunde liegenden Daten.

Besonders bei der Visualisierung eindimensionaler Daten, wie beispielsweise Mengen- oder Größenangaben durch Flächen oder Körper, sollten Datenjournalisten darauf achten, dass sich Daten und Datenvisualisierung exakt proportional zueinander verhalten. Tufte empfiehlt sogar, dass die Anzahl der Dimensionen einer Visualisierung die Anzahl der Dimensionen in den Daten nicht überschreiten sollte<sup>204</sup>. Allerdings darf diese Empfehlung missachtet werden, da sie Datenvisualisierungstechniken wie zum Beispiel *Treemaps* (s. Abb. 10) ausschließen würde, die für die Visualisierung hierarchischer Strukturen sehr gut geeignet sind.

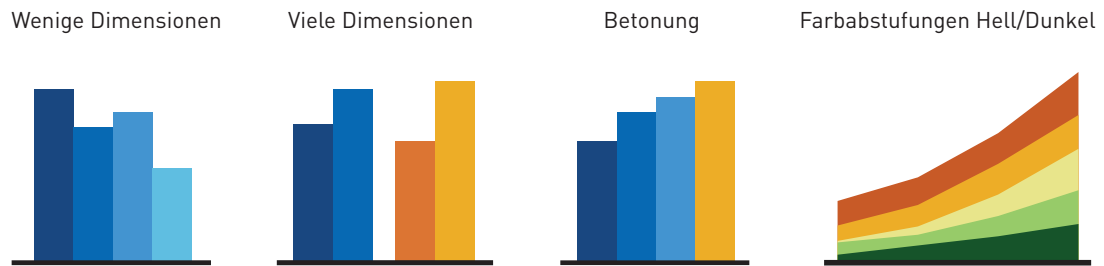
201 Vgl. Tufte (1983): The Visual Display of Quantitative Information, S. 96

202 Vgl. ebenda, S. 56 f.

203 Vgl. ebenda

204 Vgl. ebenda, S. 71



**Abbildung 11:**

Zusammenfassung der Farb-Grundsätze

**Quelle:**Nach Wong, *The Wall Street Journal guide to information graphics*

### 5.1.4 Schrift und Beschriftung

Typografie spielt in Datenvisualisierungen keine tragende Rolle. Trotzdem ist sie ein wichtiger Bestandteil guter Visualisierungen, wenn wichtige Details oder Zusammenhänge in Daten hervorgehoben oder Skalen und grafische Elemente eindeutig beschrieben werden sollen<sup>210</sup>. Damit Schrift nicht zu sehr in den Fokus einer Visualisierung rückt, sollte sie sparsam und Schriftauszeichnungen nur zur Betonung wichtiger Daten eingesetzt werden<sup>211</sup>. Ein guter Einsatz von Schrift in einer Datenvisualisierung zeigt das Projekt *Budget Forecasts* der *New York Times* (s. Kapitel 3.1.3).

## 5.2 What They Know vom Wall Street Journal

Um herauszufinden, in welchem Umfang Websites und Apps ihre Besucher ausspionieren und an wen die gesammelten Daten weitergeleitet werden, initiierte das *Wall Street Journal* (WSJ) 2010 das Projekt *What They Know*. Die dreiteilige Serie untersuchte dabei das Datensammelverhalten der fünfzig meist besuchten Websites in den USA (plus die Website des WSJ selbst)<sup>212</sup>, der fünfzig meist besuchten Kinder-Websites<sup>213</sup> und der 101 beliebtesten Apps für das *iPhone* und *Android*<sup>214</sup> auf den Einsatz von Späh-technologien und an welche Unternehmen die erspähten Daten weitergeleitet werden.

Die Ergebnisse visualisierte das WSJ in drei separaten Datenvisualisierungen. Auftakt der Serie war die Visualisierung der fünfzig US-Top-Websites<sup>215</sup> (s. Abb. 12), die in diesem Kapitel auf die Grundsätze guter Datenvisualisierungen untersucht wird.

<sup>210</sup> Vgl. Tufte (1983): *The Visual Display of Quantitative Information*, S. 56

<sup>211</sup> Vgl. Wong (2010): *The Wall Street Journal Guide to information graphics*, S. 33

<sup>212</sup> The Wall Street Journal (o.J.): *What They Know*, 18. Juli 2011

<sup>213</sup> ebenda, s. „KIDS“

<sup>214</sup> ebenda, s. „APPS“

<sup>215</sup> Vgl. The Wall Street Journal (2010): *Tracking the Trackers: Our Method*, 18. Juli 2011

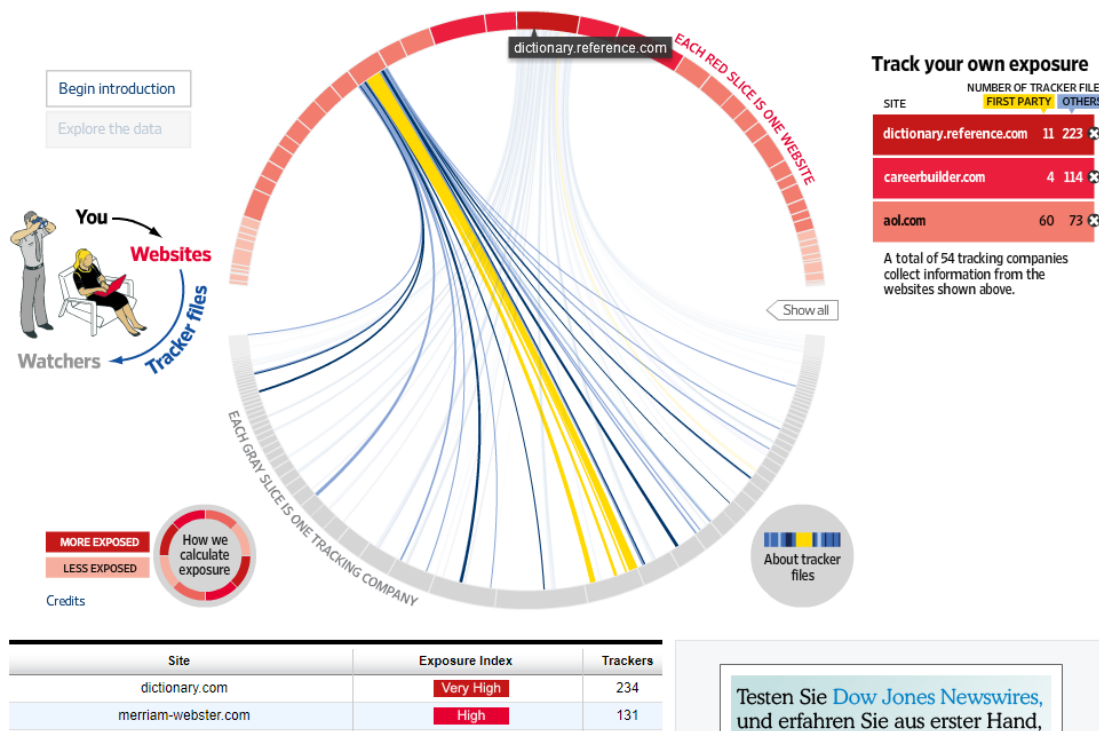


Abbildung 12:

Interaktive Datenvisualisierung  
 „What They Know“ des Wall Street Journal

Quelle:

The Wall Street Journal, What They Know

## 5.2.1 Allgemein

Auf der Startseite der Anwendung befindet sich ein Einführungstext, der den Hintergrund und den Umfang des Projekts erklärt. Parallel dazu baut sich die Anwendung langsam auf und eine erklärende Animation beschreibt den Prozess und die einzelnen Elemente der Datenvisualisierung. Da sich die Animation beliebig wiederholen lässt, wird dem Nutzer eine jederzeit verfügbare Orientierungshilfe für den Umgang mit der Visualisierung angeboten.

Sich öffnende Hinweisboxen unter der Datenvisualisierung erklären die verschiedenen Tracking-Technologien (rechter Kreis) und die Einteilung der Websites nach dem „*Exposure Index*“ (linker Kreis). Dieser findet sich als Tabelle unter den Hinweisboxen und listet die untersuchten Websites nach ihrer Punktzahl, die sie auf der Grundlage von acht Datenschutz-Kriterien erreichen<sup>216</sup>.

Die zentrale Datenvisualisierung bietet ein sehr hohes *Data-ink*-Verhältnis, da bis auf die Beschriftung der roten und blauen Kreissegmente und dem „*Show all*“-Button die gesamte Fläche der Grafik Daten entspricht und die Beschriftung der Kreissegmente sich nur zeigt, wenn der Nutzer den Mauszeiger darüber bewegt.

Die Breite der Kreissegmente im Bereich der untersuchten Websites verhält sich proportional zur Anzahl der vom *WSJ* festgestellten Tracking-Dateien bzw. im Bereich der Unternehmen zur Anzahl der empfangenen Daten. Damit stimmt die visuelle Repräsen-

<sup>216</sup> Vgl. The Wall Street Journal (2010): Tracking the Trackers: Our Method, 18. Juli 2011

tation der Datenvisualisierung mit den Daten überein und ergibt einen *Lie Factor* von 1,0. Ein besonderes Zeichen der Transparenz ist die Tatsache, dass die Datenjournalisten des *WSJ* ihre eigene Website auf Spähprogramme untersuchten und die Daten in die Visualisierung integrierten. Damit beweist die Zeitung viel Mut und riskiert unnötige Kritik, denn wie die weiter unten auf der Webseite veröffentlichte Umfrage zeigt, reagieren fast 60 Prozent der Befragten sehr besorgt auf das Thema<sup>217</sup>.

### 5.2.2 Analyse der Datenvisualisierung

In der zentralen Datenvisualisierung von *What They Know* werden die Daten in kategorisierten Kreissegmenten dargestellt, die in die einzelnen Websites (rot) und Unternehmen (blau) unterteilt sind. Abhängig von der Anzahl der verwendeten Tracker wurde eine Website einer von vier Kategorien zugewiesen und farblich abgestuft visualisiert. Die Verbindungen zwischen den Daten sammelnden Websites und den Unternehmen, die die Daten erhalten, sind durch farbige Linien visualisiert. Jede Farbe entspricht dabei einer der vier Tracking-Technologien *First-Party Tracker Files*, *Cookies*, *Bacons*, *Flash cookies*. Eine nutzerfreundliche Funktion der Datenvisualisierung ist die transparente Vorschau-Funktion auf die Verbindungen anderer Websites beim Herüberfahren des Mauszeigers über ein Segment.

Im Bereich der Unternehmen wurden die verschiedenen Tracker in untergeordneten Kreissegmenten ihren Unternehmen zugeordnet. Um die Visualisierung nicht zu überladen wurden die Datenschutzerklärungen der Tracking-Programme in ein extra Fenster ausgelagert, was sich bei Klick auf das jeweilige Programm öffnet. Die Inhalte der Bestimmungen wurden gut sichtbar mit Rot und Grün bewertet, um dem Nutzer zu zeigen, welche Bestimmungen sich positiv und welche sich negativ auf seine Privatsphäre auswirken. Damit optisch ein Unterschied für benachteiligte Menschen mit Rot-Grün-Schwäche sichtbar wird, entsteht zwischen dem leuchtstarken Rotton und dem leuchtarmen Grünton ein deutlicher Kontrast. Schrift wird in der Datenvisualisierung wenig verwendet und nur an den notwendigen Stellen. Auszeichnungen betonen dabei wichtige Informationen.

### 5.2.3 Interaktionsmöglichkeiten

Entsprechend dem Auszug „*details on demand*“ aus Daniel Keims Visual Analytics-Mantra (s. Kapitel 4.4) bietet die Anwendung des *WSJ* dem Nutzer die Möglichkeit sich die einzelnen Tracker einer Website zu den verschiedenen Unternehmen entweder in der zentralen Datenvisualisierung oder separat auf einer übersichtlicheren Unterseite anzeigen zu lassen.

Gleichzeitig können Nutzer in der zentralen Datenvisualisierung sich mehrere Websites auswählen und links neben der Visualisierung eine Liste anzeigen lassen, wie viele Tracker und Unternehmen ihre Internetaktivitäten aufzeichnen.

---

217 Wall Street Journal (o.J.): What They Know, 18. Juli 2011

## 5.2.4 Bewertung

Das Projekt *What They Know* des *Wall Street Journal* ist ein gutes Beispiel einer interaktiven Datenvisualisierung im Datenjournalismus, die die durch die Daten gewonnenen Erkenntnisse gut an den Nutzer weitergibt und sich an den Grundsätzen guter Datenvisualisierungen orientiert. Neben einer nützlichen Einführung in das Thema und sinnvollen Hinweisen bietet es sowohl einen übersichtlichen Gesamteindruck über das Trackingverhalten der fünfzig untersuchten Websites als auch einen aufschlussreichen Blick auf die Details und Tracker jeder Website und jedes Unternehmens.

Der einzige, aber deutliche Kritikpunkt des Projekts ist jedoch, dass das *Journal* die Rohdaten des Projekts nicht veröffentlicht und bereitstellt, was Nutzer daran hindert die aggregierten Daten zum Nachprüfen einzusehen oder für eigene Projekte zu nutzen.



## 6 Schlussbetrachtung

### 6.1 Zusammenfassung der Erkenntnisse

Allgemein eröffnet der Einsatz von Daten im Journalismus den Medien die Möglichkeit die vernetzte und globalisierte Gesellschaft mithilfe komplexer Daten besser zu erklären. Wie gezeigt wurde, erlaubt die Verwendung von Daten in journalistischen Berichterstattungen dabei die Erschließung neuer Zusammenhänge und die Aufdeckung unbekannter Informationen in Daten sowie eine genauere Abbildung der wirtschaftlichen, politischen und gesellschaftlichen Wirklichkeit.

Im Speziellen erreichen Medien diese Verbesserung durch die Integration von Daten auf den verschiedenen Ebenen der journalistischen Arbeit. So profitieren Datenjournalisten von kürzeren Recherchezeiten durch die Verfügbarkeit strukturierter und thematisch relevanter Daten in ihren eigenen Medienarchiven oder externen Websites und Portalen.

Weiter lassen sich Berichte über besonders datenintensive Prozesse durch den Einsatz aussagekräftiger Datenvisualisierungen statt erklärender Texte besser kommunizieren und ermöglichen Nutzern eine beschleunigte Rezeption des von Journalisten aus den Daten extrahierten Wissens.

Die Integration der Nutzer und deren Arbeitskraft in die Datenanalyse über Crowdsourcing erhöht außerdem die Bindung zwischen Medien und Nutzern, während zugleich die Angabe von Quellen und die Herausgabe von Rohdaten die Transparenz der Medien und das Vertrauen der Nutzer in die Medien stärkt.

Neue Darstellungsformen interaktiver Datenvisualisierungen, die beispielsweise als Karten oder mit Artikeln, Fotos, Videos bestückt werden, sorgen zudem für eine dichtere Informationsatmosphäre und somit für ein intensiveres Storytelling.

Um diese Potentiale von Daten im Journalismus nutzen zu können, stehen Journalisten sowie Medien allerdings vor deutlichen Herausforderungen. Journalisten müssen datensicher werden. Von der Aggregation, über die Filterung, Aufbereitung und Analyse von Daten bis zur Präsentation der gewonnenen Erkenntnisse in interaktiven Datenvisualisierungen sind sie in interdisziplinäre Teilprozesse eingebunden, die außerhalb der bekannten Anforderungen an Journalisten ansetzen. Vor allem umfangreiches Methodenwissen zur effizienten Durchführung von Datenprojekten zeichnet einen guten Datenjournalisten aus.

Medien hingegen sehen sich vor allem mit dem Aufbau von Datenteams konfrontiert, die es ihnen ermöglichen sollten unabhängig eigene Datenprojekte durchführen zu können. In diesen sollte neben der journalistischen Kompetenz auch Programmierwissen und Designwissen vorhanden sein.

Die dafür benötigten Ressourcen in Datenkompetenz und Fachpersonal sind aktuell allerdings erst in den Medienhäusern einiger weniger Pioniere vorhanden. Jedoch zeigen die steigende Anerkennung, Wahrnehmung und der zunehmende Einsatz datenjournalistischer



Inhalte in Online-Medien, dass die Notwendigkeit für das Thema von vielen Medientreibern erkannt worden ist.

Gleichzeitig steht den Entwicklungen der Meinungsmedien im Datenjournalismus ein immer größer werdender Kreis an freien Journalisten, Informationsdesignern und Programmierern gegenüber. Deren unabhängige und investigative Arbeiten mit Daten bauen einen starken Gegenpol zu dem bestehenden Angebot der Leitmedien auf und weisen auf deren drohenden Kompetenzverlust und sinkenden Einfluss hin.

Medien sollten diese Entwicklungen jedoch als Chance begreifen, um aus ihren alten Rollenbildern auszubrechen und ihren Lesern gegenüber mit mehr Vertrauen und Transparenz in der Berichterstattung zu begegnen.

Der britische *Guardian* ist sich diesem wandelnden Rollenbild des Journalismus bewusst geworden und investiert in den letzten Jahren viel in den Ausbau seiner Transparenz und Offenheit gegenüber seinen Lesern. Der *Guardian* nimmt damit eine Vorbild-Funktion für den weltweiten Datenjournalismus ein, der weitere Medien folgen sollten.

Abschließend lässt sich der Einfluss von Daten auf den Journalismus als weitreichend und das Potential von Daten als enorm zusammenfassen, nicht nur für Medien und Journalisten, sondern auch für die Nutzer und die allgemeine Öffentlichkeit.

## 6.2 Ausblick auf den Datenjournalismus

Nach der in *Abbildung 1* dargestellten Einschätzung auf das weltweite Datenwachstum bis 2015 wird die Masse an Daten und somit auch die Masse verfügbarer Daten deutlich ansteigen. Dadurch ist zu erwarten, dass Datenjournalisten in ihrer Funktion als Datenberichtersteller zunehmend wichtiger werden.

Durch die verstärkte Vernetzung der Welt und die steigende Verwendung von Metadaten in Datenbanken ist zudem anzunehmen, dass auch die Komplexität der Daten ansteigen wird. Diese Entwicklung wird sowohl neue Formen von Datenvisualisierungen erforderlich machen, als auch eine stetige Weiterbildung in dem Thema auf der Seite von Medien und Journalisten, deren eingesetzte Arbeitstechniken mit den Entwicklungen Stand halten müssen.

Zudem ist davon auszugehen, dass die medial aktivere Gesellschaft zu einem sinkenden Einfluss der Leitmedien bei einem gleichzeitig stärkeren Dialog zwischen Gesellschaft und Medien führen wird. Wie stark sich Medien für diesen Dialog öffnen wollen und in Kollaboration mit ihren Nutzern Datenprojekte erarbeiten werden, bleibt abzuwarten.

Als strategisch sehr sinnvoll kann der Aufbau von eigenen *Datastores* in Medien nach dem Vorbild des *Guardian* gesehen werden, die mit aktuell recherchierten Daten und aufbereitetem Material aus den internen Archiven gefüllt werden können. Solche *Datastores* würden Medien einerseits eine weitere Anlaufstelle zu ihrem Angebot garantieren. Andererseits könnten Medien durch den Service der Bereitstellung und Nutzung von sehr komplexen oder hyperlokalen Daten eine neue Einnahmequelle erschließen. Medien würden damit eine naheliegende Weiterentwicklung von Nachrichten Anbietern zu Datenanbietern vollziehen. Bei entsprechenden Ressourcen ist auch der Aufbau einer eigenen Datenagentur denkbar,

die auf Anfrage – zum Beispiel aus der Bevölkerung, von Institutionen oder anderen Medien – entsprechende Daten sammelt, aufbereitet und bereitstellt. Das kann weiter auch die Erstellung von geeigneten Werkzeugen zur Analyse oder interaktiven Datenvisualisierungen zur Darstellung von Daten umfassen. Die Kompetenzen für so eine Datenagentur wären in einem für den Datenjournalismus gerüsteten Medienhaus zumindest vorhanden.

## 6.3 Weiterführende Untersuchungen

Im Verlauf der Arbeit ergaben sich verschiedene Fragestellungen im thematischen Umfeld des Datenjournalismus, die aufgrund ihres Umfangs nicht näher in der Arbeit untersucht werden konnten. Die Fragestellungen eignen sich jedoch durch ihre Relevanz und Nähe zum Datenjournalismus als Gegenstand weiterer wissenschaftlicher Untersuchungen, die auf den Erkenntnissen dieser Arbeit aufbauen können:

### **Aufbau eines Data-Warehouse in Nachrichtenredaktionen**

Zur zentralen Lagerung von Datenbeständen werden sogenannte *Data-Warehouses* benutzt. Für einen funktionierenden Datenjournalismus ist die schnelle und umfangreiche Verfügbarkeit von Daten in Redaktionen essentiell wichtig. Wie die Daten von Medienhäusern in eigenen Data-Warehouses gespeichert und wie sie in den Workflow von Datenjournalisten sinnvoll integriert werden können, ist ein möglicher Ansatz für weitere wissenschaftliche Untersuchungen.

### **Opinion Mining in sozialen Netzwerken**

Neben dem in dieser Arbeit vorgestellten Scraping und Data Mining auf Websites und in Datenbanken existiert noch das sogenannte *Opinion Mining* als Sonderform des Data Mining. Häufig im Marketing zur Erforschung von Nutzerinteressen eingesetzt, kann Opinion Mining auch im Datenjournalismus eine wichtige Rolle spielen. Einsatzgebiete sind hier zum Beispiel die Gewinnung von Informationen über die Meinung der Öffentlichkeit zu bestimmten Themen als Ersatz oder Teilersatz für die häufig in Medien durchgeführten Umfragen. Durch die offenen Schnittstellen sozialer Netzwerke, über die Daten aus Netzwerken ausgelesen werden können, sind diese besonders zur Datenerhebung geeignet. Eine weiterführende wissenschaftliche Arbeit kann untersuchen, inwieweit Opinion Mining im Journalismus zur Wiedergabe der öffentlichen Meinung relevant wäre und wie die Daten sozialer Netzwerke dabei effektiv genutzt werden können.

### **Workflow und Umsetzung dynamischer Datenvisualisierungen**

Neben Daten-Visualisierungen aus Datenbanken sind auch Visualisierungen gefilterter und ständig nachströmender Daten aus sozialen Netzwerken möglich. Solche dynamischen Visualisierungen und Projekte erfordern einen angepassten Workflow, der im Rahmen einer eigenständigen Arbeit entwickelt werden kann. Darin stünden auch technische Fragen im Vordergrund, die die effektive Verarbeitung der Daten untersuchen würden.

# Literaturverzeichnis

## A

---

*Alan, Alasdair/Warden, Pete* (2011): Got an iPhone or 3G iPad? Apple is recording your moves,

URL: <http://radar.oreilly.com/2011/04/apple-location-tracking.html>,

Stand: 18. Juli 2011.

*Alexa* (o.J.), Top Sites,

URL: <http://www.alexa.com/topsites/global>,

Stand: 18. Juni 2011.

## B

---

*Bauer Media Akademie* (2007): Zukunftswerkstatt Medien: MedienExpertenPanel,

URL: [http://www.bauermedia.de/uploads/media/Medienexperten-](http://www.bauermedia.de/uploads/media/Medienexperten-Panel_1_-_2007_01.pdf)

[Panel\\_1\\_-\\_2007\\_01.pdf](http://www.bauermedia.de/uploads/media/Medienexperten-Panel_1_-_2007_01.pdf),

Stand: 14. Juli 2011.

*Beckedahl, Markus* (2011): Datenjournalismus und die Zukunft der Berichterstattung,

URL: <http://netzpolitik.org/2011/gastbeitrag-datenjournalismus-und-die-zukunft-der-berichterstattung/>,

Stand: 14. Juli 2011.

*berlin.de* (o.J.): Open Cities,

URL: <http://www.berlin.de/projektzukunft/networking/internationale-kooperationen/open-cities/>,

Stand: 19. Juni 2011.

*Berners-Lee, Timothy* (2007): Hearing on the „Digital Future of the United States: Part I -- The Future of the World Wide Web“,

URL: <http://dig.csail.mit.edu/2007/03/01-ushouse-future-of-the-web.pdf>,

Stand: 19. Juni 2011.

*Berners-Lee, Timothy* (2010): Open, Linked Data for a Global Community,

URL: <http://www.youtube.com/watch?v=ga1aSJXCFe0>,

Stand: 18. Juli 2011

*Biermann, Kai* (2011): Was Vorratsdaten über uns verraten,

URL: <http://www.zeit.de/digital/datenschutz/2011-02/vorratsdaten-malte-spitz>,

Stand: 14. Juli 2011.

*BITKOM* (2011): Information Overload? Wie die Deutschen mit Nachrichtenflut und Medienvielfalt umgehen, URL: [http://www.bitkom.org/files/documents/BITKOM\\_Praesentation\\_Info-Management\\_31\\_03\\_2011%281%29.pdf](http://www.bitkom.org/files/documents/BITKOM_Praesentation_Info-Management_31_03_2011%281%29.pdf), Stand: 18. Juni 2011.

*Bizer, Chris et al.* (o.J.): How to Publish Linked Data on the Web, The RDF Data Model, URL: <http://www4.wiwi.fu-berlin.de/bizer/pub/LinkedDataTutorial/#datamodel>, Stand: 18. Juli 2011.

*Bradshaw, Paul* (2010): How to be a data journalist, URL: <http://www.guardian.co.uk/news/datablog/2010/oct/01/data-journalism-how-to-guide>, Stand: 14. Juli 2011.

*Brochardt, Mirka* (2011): Brecht das Informationsmonopol!, URL: [http://blogs.taz.de/hausblog/2011/04/09/brecht\\_das\\_informationsmonopol/](http://blogs.taz.de/hausblog/2011/04/09/brecht_das_informationsmonopol/), Stand: 18. Juli 2011.

*Bundesministerium für Justiz* (2006): §1 Abs.1 des Informationsweiterverwendungsgesetz, URL: <http://www.gesetze-im-internet.de/iwg/BJNR291300006.html>, Stand: 19. Juni 2011.

*Bundesministerium für Wirtschaft und Technologie* (2010): Dresdner Vereinbarung, URL: <http://www.bmwi.de/BMWi/Redaktion/PDF/Publikationen/Technologie-und-Innovation/it-gipfel-2010-dresdner-vereinbarung,property=pdf,bereich=bmwi,sprache=de,rwb=true.pdf>, Stand: 19. Juni 2011.

## C

---

*Chivers, Christopher J. et al.* (2010): View is Bleaker Than Official Portrayal of War in Afghanistan, URL: <http://www.nytimes.com/2010/07/26/world/asia/26warlogs.html>, Stand: 15. Juli 2011.

*Cisco Visual Networking Index* (2011): White Paper, URL: [http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white\\_paper\\_c11-481360.pdf](http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf), Stand: 18. Juni 2011.

*Coates, Sam* (2009): Behind the Story: MPs turn fire on Andrew Walker, man who approved claims, URL: <http://www.timesonline.co.uk/tol/news/politics/article6289143.ece>, Stand: 14. Juli 2011.

*Conlan, Tara* (2006): BBC to pay for viewers' pictures, URL: <http://www.guardian.co.uk/media/2006/nov/16/broadcasting.bbc>, Stand: 18. Juli 2011.

*Creative Commons* (o.J.): Was ist CC?,  
URL: <http://de.creativecommons.org/was-ist-cc/>,  
Stand: 19. Juni 2011.

*Creative Commons* (o.J.): Mehr über die Lizenzen,  
URL: <http://creativecommons.org/licenses/>,  
Stand: 19. Juni 2011.

## D

---

*Dambeck, Thorsten* (2004): Aufstand gegen die Hüter des Wissens,  
URL: <http://www.spiegel.de/wissenschaft/mensch/0,1518,313112,00.html>,  
Stand: 18. Juli 2011.

*Darnstädt, Thomas* (2007): Die Gratis-Kultur,  
URL: <http://www.spiegel.de/spiegelspecial/0,1518,490777,00.html>,  
Stand: 18. Juni 2011.

*Data.gov* (o.J.): Data Policy,  
URL: <http://www.data.gov/datapolicy/>,  
Stand: 19. Juni 2011.

*data.gov.uk* (o.J.): What is our approach?,  
URL: <http://data.gov.uk/project#q2>,  
Stand: 18. Juli 2011.

*data.gov.uk* (o.J.): How were the datasets in Data.gov selected?,  
URL: <http://data.gov.uk/faq#q9>,  
Stand: 18. Juli 2011.

*DBpedia* (o.J.):  
URL: <http://bit.ly/rdfwolkenkratzer>,  
Stand: 18. Juli 2011.

*Deutscher Presserat* (o.J.): Pressekodex Ziffer 2 - Sorgfalt,  
URL: <http://www.presserat.info/inhalt/der-pressekodex/pressekodex/richtlinien-zu-ziffer-2.html>,  
Stand: 14. Juli 2011.

*Deutsches Institut für Wirtschaftsforschung* (2003): Internetnutzung in Deutschland: Nach Boom nun langsamerer Anstieg erwartet,  
URL: [http://www.diw.de/de/diw\\_01.c.10370.de/themen\\_nachrichten/internetnutzung\\_in\\_deutschland\\_nach\\_boom\\_nun\\_langsamerer\\_anstieg\\_erwartet.html](http://www.diw.de/de/diw_01.c.10370.de/themen_nachrichten/internetnutzung_in_deutschland_nach_boom_nun_langsamerer_anstieg_erwartet.html),  
Stand: 18. Juni 2011.

---

## E

---

*Eberle, Matthias et al.* (2010): Das Ende der Gratiskultur im Internet ist angekommen,  
URL: <http://www.handelsblatt.com/unternehmen/it-medien/das-ende-der-gratiskultur-im-internet-ist-gekommen/3631270.html>,  
Stand: 18. Juni 2011.

*Elmert, Christina* (2011): Datenjournalismus ganz praktisch,  
URL: <http://recherche-info.de/2011/04/17/datenjournalismus-ganz-praktisch-christina-elmer-dpa-auf-der-republica/>,  
Stand: 15. Juli 2011.

*Empson, Rip* (2011): When It Comes To New Journalism, Transparency Is the New Objectivity,  
URL: <http://techcrunch.com/2011/05/23/jeff-jarvis-when-it-comes-to-new-journalism-transparency-is-the-new-objectivity/>,  
Stand: 14. Juli 2011.

*Engadget* (o.J.): About Us,  
URL: <http://www.engadget.com/about/>,  
Stand: 18. Juli 2011.

*Eurostat* (o.J.): Rechtlicher Hinweis,  
URL: [http://ec.europa.eu/geninfo/legal\\_notices\\_de.htm](http://ec.europa.eu/geninfo/legal_notices_de.htm),  
Stand: 19. Juni 2011.

---

## F

---

*Facebook* (o.J.): Statistics,  
URL: <http://www.facebook.com/press/info.php?statistics>,  
Stand: 18. Juli 2011.

*Fayyad, Usama et al.* (1996): From Data Mining to Knowledge Discovery in Databases,  
in: *AI Magazine*, Nr. 3, S. 37-54.

*Fischer, Sebastian et al.* (2011): Copy, paste, delete;  
URL: <http://www.spiegel.de/politik/deutschland/0,1518,746260,00.html>;  
Stand: 14. Juli 2011.

*Flickr* (o.J.): How can I copyright my photos?,  
URL: <http://www.flickr.com/help/photos/#87>,  
Stand: 19. Juni 2011.

---

**G**

---

*Georgi, Oliver* (2011): Anfang von F.A.Z. abgeschrieben,  
URL: <http://www.faz.net/artikel/C30923/guttenberg-dissertation-anfang-bei-f-a-z-abgeschrieben-30327825.html>,  
Stand: 14. Juli 2011.

*Grimberg, Steffen* (2009): „Das alte System ist kaputt“,  
URL: <http://www.taz.de/!33828/>,  
Stand: 19. Juni 2011.

*Grimme Online Award* (2011): Begründung der Jury [Verräterisches Handy]  
URL: <http://www.grimme-institut.de/html/index.php?id=1345>,  
Stand: 14. Juli 2011.

*Grimme Online Award* (2011): Preisträger 2011,  
URL: <http://www.grimme-institut.de/html/index.php?id=1122>,  
Stand: 14. Juli 2011.

*Grimme Online Award* (2011): Begründung der Jury [Guttenplag Wiki]  
URL: <http://www.grimme-institut.de/html/index.php?id=1344>,  
Stand: 14. Juli 2011.

*Grossmann, Lev* (1998): New Free License to Cover Content Online,  
Archivierungs-URL: <http://web.archive.org/web/20000619122406/http://www.time.com/time/digital/daily/0,2822,621,00.html>,  
Stand: 18. Juli 2011.

*guardian.co.uk* (2009): Investigate your MP's expenses,  
URL: <http://mps-expenses.guardian.co.uk/>,  
Stand: 14. Juli 2011.

*guardian.co.uk* (o.J.): Data blogosphere: today's picks,  
URL: <http://www.guardian.co.uk/news/datablog/blogosphere>,  
Stand: 14. Juli 2011.

*guardian.co.uk* (o.J.): Datablog,  
URL: <http://www.guardian.co.uk/news/datablog>,  
Stand: 14. Juli 2011.

*Gunter, Joel* (2011): Newspaper Awards 2011: Guardian takes home digital prizes for websites and datablog,  
URL: <http://www.journalism.co.uk/news/newspaper-awards-2011-guardian-takes-home-digital-prizes-for-website-and-datablog/s2/a544319/>,  
Stand: 14. Juli 2011.



*GuttenPlag* (o.J.):

URL: [http://de.guttenplag.wikia.com/wiki/GuttenPlag\\_Wiki](http://de.guttenplag.wikia.com/wiki/GuttenPlag_Wiki),  
Stand: 14. Juli 2011.

## H

---

*H2 media factory* (o.J.): Rechtliche Hinweise,

URL: <http://www.h2.de/rechtliche-hinweise>,  
Stand: 18. Juli 2011.

*Holovaty, Adrian* (2006): A fundamental way newspaper sites need to change,

URL: <http://www.holovaty.com/writing/fundamental-change/>,  
Stand: 14. Juli 2011.

## J

---

*Jansen, Frank et al.* (2010): Eine furchtbare Bilanz,

URL: <http://www.zeit.de/2010/38/Rechte-Gewalt>,  
Stand: 15. Juli 2011.

*Juskalian, Russ* (2008): Interview with Clay Shirky, Part I,

URL: [http://www.cjr.org/overload/interview\\_with\\_clay\\_shirky\\_par.php?page=all](http://www.cjr.org/overload/interview_with_clay_shirky_par.php?page=all),  
Stand: 18. Juni 2011.

## K

---

*Keese, Christoph* (2007): Willkommen bei WELT ONLINE,

URL: [http://www.welt.de/personalisierung/article698156/Willkommen\\_bei\\_WELT\\_ONLINE.html](http://www.welt.de/personalisierung/article698156/Willkommen_bei_WELT_ONLINE.html),  
Stand: 18. Juli 2011.

*Keim, Daniel et al.* (2010): Mastering the Information Age. Solving Problems with Visual Analytics, Bad Langensalza.

*Keim, Daniel et al.* (2006): Challenges in Visual Data Analysis,

in: Information Visualization, Nr. 4, S. 9-16,  
URL: <http://kops.ub.uni-konstanz.de/handle/urn:nbn:de:bsz:352-opus-69128>,  
Stand: 14. Juli 2011.

*Keim, Daniel et al.* (2008): Visual Analytics: Definition, Process, and Challenges, Information, Visualization;

URL: [http://hal-lirmm.ccsd.cnrs.fr/docs/00/27/27/79/PDF/VACchapter\\_final.pdf](http://hal-lirmm.ccsd.cnrs.fr/docs/00/27/27/79/PDF/VACchapter_final.pdf),  
Stand: 15. Juli 2011.

*Keim, Daniel* (2005): Scaling Visual Analytics to Very Large Data Sets,  
URL: <http://infovis.uni-konstanz.de/events/VisAnalyticsWs05/pdf/03DanielKeim.pdf>,  
Stand: 15. Juli 2011.

*Keim, Daniel A.* (2002): Information Visualization and Visual Data Mining,  
in: TVCG, 7. Jg., Nr. 1, S. 100-107,  
URL: <http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.87.1208>,  
Stand: 15. Juli 2011.

*Klaß, Christian* (2011): Heimat von über 50 Millionen Blogs,  
URL: <http://www.golem.de/1107/84844.html>,  
Stand: 18. Juli 2011.

*Knieper, Thomas* (1995): Infographiken: Das visuelle Informationspotential der Tageszeitung, München.

*Kohlhepp, Björn* (2011): Twitter feiert fünften Geburtstag,  
URL: <http://www.mainpost.de/ueberregional/wirtschaft/mainpostwirtschaft/Twitter-feiert-fuenften-Geburtstag;art9485,6241021>,  
Stand: 18. Juli 2011.

*Kolo, Castulus/Meyer-Lucht, Robin* (2007): Erosion der Intensivleserschaft,  
in: M&K, 55. Jg., Nr. 4, S. 513-533.

*Kramer, Staci D.* (2007): New York Times To Close TimesSelect Effective Midnight Tuesday; Will Open Last 20 Years Of Archives;  
URL: <http://paidcontent.org/article/419-new-york-times-to-close-timesselect-effective-wednesday/>;  
Stand: 18. Juni 2011.

## L

---

*von La Roche, Walther* (1975): Einführung in den praktischen Journalismus,  
18. Auflage, Berlin.

*Langner, Ulrike* (2011): Daten rücken in den Fokus,  
in: Fachjournalist, Nr. 2, S. 4-9.

*Langner, Ulrike* (2011): Schaubild statt Klickstrecke,  
in: Medium Magazin, Nr. 1-2, S. 40-41.

*Langner, Ulrike* (2011): „Man macht sich überprüfbar“,  
in: Medium Magazin, Nr. 1-2, S. 43-44.

*Langner, Ulrike* (o.J.): Thema Datenjournalismus: Das Beispiel „Zeit online“,  
URL: <http://www.mediummagazin.de/magazin-plus/thema-datenjournalismus-das-beispiel-zeit-online/>,  
Stand: 15. Juli 2011.

*Linked Open Data* (o.J.): Why do you think this project will have a wide impact?,  
URL: <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData#FAQ>,  
Stand: 19. Juni 2011.

*Lipowicz, Alice* (2011): More federal agencies blog on Tumblr,  
URL: <http://cw.com/articles/2011/07/06/more-federal-agencies-blogging-on-tumblr.aspx>,  
Stand: 18. Juli 2011.

*Lischka, Konrad* (2011): „New York Times“ führt das Digitalabo ein,  
URL: <http://www.spiegel.de/netzwelt/web/0,1518,751555,00.html>,  
Stand: 19. Juni 2011.

*Littledale, Andy/Baily, Gavin* (2010): Guardian News Traces,  
URL: <http://www.guardian.co.uk/open-platform/blog/guardian-news-traces>,  
Stand: 14. Juli 2011.

*von Lucke, Jörn/Geiger, Christian* (2010): Open Government Data: Frei verfügbare Daten des öffentlichen Sektors,  
URL: <http://www.zeppelin-university.de/deutsch/lehrstuehle/ticc/TICC-101203-OpenGovernmentData-V1.pdf>,  
Stand: 19. Juni 2011.

## M

---

*Matzat, Lorenz* (2011): München will seinen Datenschutz heben,  
URL: <http://blog.zeit.de/open-data/2011/01/26/muenchen-open-government/>,  
Stand: 19. Juni 2011.

*Matzat, Lorenz* (2011): Daten sind der Treibstoff des Journalismus,  
URL: <http://www.datenjournalist.de/daten-sind-der-treibstoff-des-journalismus/>,  
Stand: 14. Juli 2011.

*Matzat, Lorenz* (2011) Fluglärmkarte BBI: Making of,  
URL: <http://www.datenjournalist.de/fluglarmkarte-bbi-making-off/>,  
Stand: 14. Juli 2011.

*Max-Planck-Gesellschaft* (o.J.): Berliner Erklärung,  
URL: <http://oa.mpg.de/lang/de/berlin-prozess/berliner-erklarung/>,  
Stand: 19. Juni 2011.

*Mazetti, Mark et al.* (2010): Pakistan Aids Insurgency in Afghanistan, Reports Assert,  
URL: <http://www.nytimes.com/2010/07/26/world/asia/26isi.html>,  
Stand: 14. Juli 2011.

*McLean, Alan* (2010): Data Driven Journalism - Telling Stories Online,  
URL: <http://www.slideshare.net/amclean/data-driven-journalism-telling-stories-online>,  
Stand: 14. Juli 2011.

*Meckel, Miriam* (2006): Was vom Tage übrig bleibt,  
URL: <http://www.miriammeckel.de/2006/11/29/was-vom-tage-uebrig-bleibt/>,  
Stand: 19. Juni 2011.

*Meckel, Miriam* (2011): Journalisten an der Crowdsourcing-Front,  
URL: [http://www.focus.de/digital/internet/dld-2011/debate/tid-20968/medienwandel-journalisten-an-der-crowdsourcing-front\\_aid\\_589439.html](http://www.focus.de/digital/internet/dld-2011/debate/tid-20968/medienwandel-journalisten-an-der-crowdsourcing-front_aid_589439.html),  
Stand: 14. juli 2011.

*Meyer-Lucht, Robin* (2003): Sinkende Auflagen, Einbrüche im Anzeigengeschäft,  
Konkurrent Internet, in: Analysen der Friedrich-Ebert-Stiftung zur  
Informationsgesellschaft, Nr. 9, S. 1

*Mruck, Katja et al.* (2004): Open Access: (Social) Sciences as Public Good,  
URL: <http://www.qualitative-research.net/index.php/fqs/article/viewArticle/624/1351>,  
Stand: 19. Juni 2011.

*Mühlberger, Herbert* (o.J.): Linking Open Data: Anwendungen für Endanwender  
URL: [http://blog.muehlburger.at/wp-content/uploads/2009/01/linking\\_open\\_data\\_anwendungen\\_fuer\\_endanwender.pdf](http://blog.muehlburger.at/wp-content/uploads/2009/01/linking_open_data_anwendungen_fuer_endanwender.pdf),  
Stand: 19. Juni 2011.

## N

---

*Neuroth, Oliver* (2011): „Schwarmintelligenz“ im Kampf gegen Plagiat,  
URL: <http://www.tagesschau.de/inland/guttenplag100.html>,  
Stand: 14. Juli 2011.

*New York Times* (2010): Budget Forecasts, Compared with Reality;  
URL: <http://www.nytimes.com/interactive/2010/02/02/us/politics/20100201-budget-porcupine-graphic.html>;  
Stand: 14. Juli 2011.

*New York Times* (2010): Piecing Together the Reports, and Deciding What to Publish,  
URL: <http://www.nytimes.com/2010/07/26/world/26editors-note.html>,  
Stand: 14. Juli 2011.

*Noack, Rick* (2011): „Wir alle schauen kino.to“,  
URL: <http://www.spiegel.de/netzwelt/netzpolitik/0,1518,767805,00.html>,  
Stand: 19. Juni 2011.

---

## O

---

*Olorode, Irene* (2010): Scraping als neue Methode zur Erhebung persönlichster Daten,  
URL: <http://futurezone.at/digitallife/1438-scraping-als-neue-methode-zur-erhebung-persoendlichster-daten.php>,  
Stand: 15. Juli 2011.

*open-access.net* (o.J.): Die Geschichte der Open-Access-Bewegung,  
URL: [http://open-access.net/de/allgemeines/was\\_bedeutet\\_open\\_access/geschichte/](http://open-access.net/de/allgemeines/was_bedeutet_open_access/geschichte/),  
Stand: 19. Juni 2011.

*open-access.net* (o.J.): Was bedeutet Open Access,  
URL: [http://open-access.net/de/allgemeines/was\\_bedeutet\\_open\\_access/](http://open-access.net/de/allgemeines/was_bedeutet_open_access/),  
Stand: 19. Juni 2011.

*OpenDataCity* (2011): Fluglärmkarte BBI,  
URL: <http://www.opendatacity.de/fluglarmkarte/>,  
Stand: 14. Juli 2011.

*OpenDataCity* (2011): Vorratsdatenspeicherung,  
URL: <http://www.opendatacity.de/vorratsdatenspeicherung/>,  
Stand: 14. Juli 2011.

*Open Data Network* (2010): Open Government Data Principles,  
URL: [http://wiki.opendata-network.org/Open\\_Government\\_Data\\_Principles](http://wiki.opendata-network.org/Open_Government_Data_Principles),  
Stand: 18. Juli 2011.

*O'Reilly, Tim* (2005): What ist Web 2.0,  
URL: <http://oreilly.com/web2/archive/what-is-web-20.html>,  
Stand: 19. Juni 2011.

*Ostrow, Adam* (2008): iReport Goes Live; CNN Gets Citizen Journalism,  
URL: <http://mashable.com/2008/02/13/cnn-ireport-launch/>,  
Stand: 18. Juli 2011.

o.V. (2010): Über zwei Milliarden Internet-Nutzer,  
URL: [http://www.rp-online.de/digitale/internet/Ueber-zwei-Milliarden-Internet-Nutzer\\_aid\\_922845.html](http://www.rp-online.de/digitale/internet/Ueber-zwei-Milliarden-Internet-Nutzer_aid_922845.html),  
Stand: 18. Juni 2011.

o.V. (2000): Aus STERN Online wird künftig die Marke stern.de,  
URL: [http://www.presseportal.de/pm/6329/109514/gruner\\_jahr\\_stern](http://www.presseportal.de/pm/6329/109514/gruner_jahr_stern),  
Stand: 18. Juli 2011.

- o. V. (2007): Studie: Tageszeitungsleser wandern ab ins Internet,  
URL: <http://it-republik.de/business-technology/news/Studie-Tageszeitungsleser-wandern-ab-ins-Internet-040117.html>,  
Stand: 18. Juni 2011.
- o. V. (2005): Wo Promis endlich reden dürfen,  
URL: <http://www.ksta.de/html/artikel/1115222913123.shtml>,  
Stand: 18. Juli 2011.
- o. V. (2011): Aigner fordert Klarheit von Apple,  
URL: <http://www.wallstreet-online.de/nachricht/3139313-aigner-fordert-klarheit-von-apple>,  
Stand: 19. Juni 2011.
- o. V. (2011): Update: iPhone soll jetzt weniger Daten sammeln,  
URL: <http://www.heute.de/ZDFheute/inhalt/31/0,3672,8237183,00.html>,  
Stand: 18. Juli 2011.
- o. V. (2009): Wikipedia Moving From GFDL To Creative Commons License,  
URL: <http://news.slashdot.org/story/09/05/21/2317253/Wikipedia-Moving-From-GFDL-To-Creative-Commons-License>,  
Stand: 19. Juni 2011.
- o. V. (2009): Expenses: How MP's expenses became a hot topic,  
URL: <http://www.telegraph.co.uk/news/newstopics/mps-expenses/5294350/Expenses-How-MPs-expenses-became-a-hot-topic.html>,  
Stand: 14. Juli 2011.
- o. V. (2007): Köhler unterzeichnet Gesetz zur Vorratsdatenspeicherung,  
URL: <http://www.heise.de/newsticker/meldung/Koehler-unterzeichnet-Gesetz-zur-Vorratsdatenspeicherung-173690.html>,  
Stand: 14. Juli 2011.
- o. V. (2011): Bei diesen Autoren könnte Guttenberg geklaut haben,  
URL: <http://www.welt.de/politik/deutschland/article12568246/Bei-diesen-Autoren-koennte-Guttenberg-geklaut-haben.html>,  
Stand: 14. Juli 2011.

## P

---

*Pérez-Peña, Richard* (2007): Times to Stop Charging for Parts of Its Web Site,  
URL: <http://www.nytimes.com/2007/09/18/business/media/18times.html>,  
Stand: 18. Juni 2011.

*Preuß, Roland/Schultz, Tanjev* (2011): Guttenberg soll bei Doktorarbeit abgeschrieben haben, URL: <http://www.sueddeutsche.de/politik/plagiatsvorwurf-gegen-verteidigungsminister-guttenberg-soll-bei-doktorarbeit-abgeschrieben-haben-1.1060774>,  
Stand: 14. Juli 2011.

*Pohlmann, Sonja* (2011): Schwarm und Schwärmer,  
URL: <http://www.tagesspiegel.de/medien/schwarm-und-schwaermer/3872214.html>,  
Stand: 14. Juli 2011.

## R

---

*Rogers, Simon* (2010): Information is power,  
URL: <http://www.guardian.co.uk/media/2010/may/24/data-journalism>,  
Stand: 19. Juni 2011.

*Rogers, Simon* (2009): How to crowdsource MP's expenses,  
URL: <http://www.guardian.co.uk/news/datablog/2009/jun/18/mps-expenses-houseofcommons>,  
Stand: 14. Juli 2011.

*Rogers, Tony* (o.J.): Why Are Newspapers Dying?,  
URL: <http://journalism.about.com/od/trends/a/dyingpapers.htm>,  
Stand: 18. Juni 2011.

*Rosenberg, Daniel* (2003): Early Modern Information Overload,  
in: Journal of the History of Ideas, 64. Jg., Nr. 1, S. 1-9.

## S

---

*Schiller, Vivian* (o.J.): A Letter to Readers About TimesSelect,  
URL: <http://www.nytimes.com/ref/membercenter/lettertoreaders.html>,  
Stand: 18. Juni 2011.

*Schmidt, Nils* (2005): Der Weg zum eigenen Blog,  
URL: <http://www.stern.de/digital/online/internet-trend-der-weg-zum-eigenen-blog-545969.html>,  
Stand: 18. Juli 2011.

*Schmitz, Daniela* (2003): Online-Stellenbörsen profitieren von Anzeigenrückgang bei Zeitungen,  
URL: <http://www.heise.de/newsticker/meldung/Online-Stellenboersen-profitieren-von-Anzeigenrueckgang-bei-Zeitungen-82079.html>,  
Stand: 18. Juli 2011.

*Schnabel, Ulrich* (2011): „Ich wollte es nicht glauben“,  
URL: <http://www.zeit.de/2011/09/Interview-Fischer-Lescano>,  
Stand: 14. Juli 2011.

*Schneider, Wolf/Raue, Paul-Josef* (1998): Das neue Handbuch des Journalismus,  
4. Auflage, Reinbeck bei Hamburg.



*Scraperwiki.com* (o.J.): Request data,

URL: [http://scraperwiki.com/request\\_data/#public](http://scraperwiki.com/request_data/#public),

Stand: 20. Juli 2011.

*Shirky, Clay* (2008): Web 2.0 Expo NY: Clay Shirky (shirky.com) It's Not Information Overload. It's Filter Failure.,

URL: <http://blip.tv/web2expo/web-2-0-expo-ny-clay-shirky-shirky-com-it-s-not-information-overload-it-s-filter-failure-1283699>,

Stand: 18. Juni 2011.

*Sixtus, Mario* (2010): Labor-Wochen beim Elektrischen Reporter,

URL: <http://blinkenlichten.com/demnaechst-labor-wochen-beim-elektrischen-reporter>,

Stand: 18. Juli 2011.

*Statistisches Bundesamt Deutschland* (o.J.): Hinweise zum Datenschutz,

URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Service/Impressum/Datenschutzerklaerung,templateId=renderPrint.psml>,

Stand: 19. Juni 2011.

*Sternberg, Josh* (2011): Why Curation Is Important to the Future of Journalism,

URL: <http://mashable.com/2011/03/10/curation-journalism/>,

Stand: 14. Juli 2011.

*Stray, Jonathan* (2010): How The Guardian is pioneering data journalism with free tools,

URL: <http://www.niemanlab.org/2010/08/how-the-guardian-is-pioneering-data-journalism-with-free-tools/>,

Stand: 14. Juli 2011.

*Stöcker, Christian* (2011): Netz besiegt Minister,

URL: <http://www.spiegel.de/netzwelt/netzpolitik/0,1518,748358,00.html>,

Stand: 14. Juli 2011.

---

## T

---

*taz.de* (2011): Parteispenden Recherche 1994-2009 - ab 10 000 Euro,

URL: <http://www.taz.de/1/politik/parteispenden-recherche/>,

Stand: 14. Juli 2011.

*taz.de* (2011): Parteispenden Watch 2009 - ab 10 000 Euro,

URL: <http://www.taz.de/1/politik/parteispenden-watch/>,

Stand: 14. Juli 2011.

*taz.de* (2011): Fluglärm-Karte BBI-Airport Berlin,

URL: <http://www.taz.de/1/berlin/fluglaerm-bbi/>,

Stand: 14. Juli 2011.

*The Free Content and Expression Definition* (o.J.): Definition,  
URL: <http://freedomdefined.org/Definition/De>,  
Stand: 19. Juni 2011.

*Toffler, Alvin* (1970): *Future Shock*,  
15th printing, New York, NY.

*Tufte, Edward R.* (1983): *The Visual Display of Quantitative Information*  
16th ed., Cheshire, CT.

*Twitter* (2011): 200 million Tweets per day,  
URL: <http://blog.twitter.com/2011/06/200-million-tweets-per-day.html>,  
Stand: 18. Juli 2011.

## V

---

*Voß, Jochen* (2007): „Zapp“ und „Extra3“ unter Creative-Commons-Lizenz,  
URL: [http://www.dwdl.de/nachrichten/13486/zapp\\_und\\_extra\\_3\\_unter\\_creativcom\\_monslizenz/](http://www.dwdl.de/nachrichten/13486/zapp_und_extra_3_unter_creativcom_monslizenz/),  
Stand: 18. Juli 2011.

## W

---

*Waeber, Matthew/Hill, Tim* (2011): Yemen, Syria and Middle East unrest,  
URL: <http://www.guardian.co.uk/world/middle-east-live/2011/jun/06/yemen-syria-middle-east-unrest>,  
Stand: 18. Juli 2011.

*The Wall Street Journal* (o.J.): What They Know,  
URL: <http://blogs.wsj.com/wtk/>,  
Stand: 18. Juli 2011.

*The Wall Street Journal* (2010): Tracking the Trackers: Our Method,  
URL: <http://online.wsj.com/article/SB10001424052748703977004575393121635952084.html>,  
Stand: 18. Juli 2011.

*Webster, Philip* (2009): Parliament's darkest day: MPs suspended and Michael Martin at risk, URL: <http://www.timesonline.co.uk/tol/news/politics/article6290054.ece>,  
Stand: 14. Juli 2011.

*Weiss, Peter* (2002): *Borders in Cyberspace: Conflicting Public Sector Information Policies and their Economic Impacts*,  
URL: [http://www.weather.gov/sp/Borders\\_report.pdf](http://www.weather.gov/sp/Borders_report.pdf),  
Stand: 19. Juni 2011.

*Wong, Dona M.* (2010): *The Wall Street Journal Guide to Information Graphics*,  
New York, NY.

*World Data Center System* (o.J.): *About the World Data Center System*,  
URL: <http://www.ngdc.noaa.gov/wdc/about.shtml>,  
Stand: 19. Juni 2011.

*World Data Center System* (o.J.): *The World Data Center System*,  
URL: <http://www.ngdc.noaa.gov/wdc/wdcmain.html>,  
Stand: 18. Juli 2011.

*Wu, Suzanne* (2011): *How Much Information Is There in the World?*,  
URL: [http://uscnews.usc.edu/science\\_technology/how\\_much\\_information\\_is\\_there\\_in\\_the\\_world.html](http://uscnews.usc.edu/science_technology/how_much_information_is_there_in_the_world.html),  
Stand: 18. Juni 2011.

## Z

---

*ZEIT ONLINE* (2011): *Verräterisches Handy*,  
URL: <http://www.zeit.de/datenschutz/malte-spitz-vorratsdaten>,  
Stand: 14. Juli 2011.

*ZEIT ONLINE* (2010): *Todesopfer rechter Gewalt 1990-2010*,  
URL: <http://www.zeit.de/gesellschaft/zeitgeschehen/todesopfer-rechter-gewalt>,  
Stand: 15. Juli 2011.

# Eigenständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe. Stellen, die wörtlich oder sinngemäß aus Quellen entnommen wurden, sind als solche kenntlich gemacht. Diese Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt.

---

Mittweida, 25.07.2011

Alexander Haase